

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra aplikované matematiky

Variační Bayesovské metody

Variational Bayesian methods

Zadání diplomové práce

Student: **Bc. Petra Kozielová**

Studijní program: N2647 Informační a komunikační technologie

Studijní obor: 1103T031 Výpočetní matematika

Téma: Variační bayesovské metody
Variational Bayesian methods

Jazyk vypracování: čeština

Zásady pro vypracování:

Bayesovská statistika představuje elegantní nástroj pro odhadování složitých statistických modelů a nachází uplatnění ve všech oblastech, kde nelze zanedbat vliv neurčitosti.

Klíčovým prvkem bayesovského přístupu je tzv. aposteriorní rozdělení, které reprezentuje celkovou informaci o neznámých parametrech modelu. Aposteriorní rozdělení kombinuje apriorní informaci (dostupnou dříve, než jsou zpracována data) s informací z dat. Toto rozdělení ale velmi často bývá natolik složité, že je nutné jej pro další výpočty aproximovat. K tomu se nejčastěji používají výpočetně náročné Markov Chain Monte Carlo (MCMC) metody, kdy je rozdělení aproximováno pomocí velkého počtu z něj generovaných vzorků. Alternativou k tomuto přístupu jsou variační bayesovské metody, kdy je aposteriorní rozdělení analyticky aproximováno jednodušším rozdělením z předem zvolené třídy. Často tak lze dosáhnout velmi dobré aproximace, přičemž následné výpočty jsou výrazně rychlejší než při využití Monte Carlo metod.

Pokyny k vypracování:

- Seznamte se se základními myšlenkami bayesovské statistiky.
- Nastudujte principy bayesovských variačních metod.
- Po dohodě se školitelem aplikujte variační bayesovské metody na zvolený statistický model (předběžný návrh: konečná lognormální směs nebo skryté markovovské modely).

Seznam doporučené odborné literatury:

Berger, James O. Statistical decision theory and Bayesian analysis. Springer Science & Business Media, 1985.

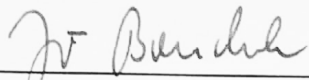
Robert, Christian. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media, 2007.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

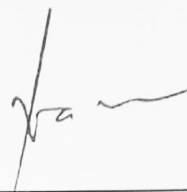
Vedoucí diplomové práce: **Ing. Jan Kracík, Ph.D.**

Datum zadání: 01.09.2017

Datum odevzdání: 30.04.2018



doc. RNDr. Jiří Bouchala, Ph.D.
vedoucí katedry



prof. Ing. Pavel Brandštetter, CSc.
děkan fakulty

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně. Uvedla jsem všechny literární
prameny a publikace, ze kterých jsem čerpala.

V Ostravě 30. dubna 2018

Kozlová
.....

Na tomto místě bych ráda poděkovala Ing. Janovi Kracíkovi, Ph.D. za cenné rady, ochotu a trpělivost při psaní této diplomové práce.

Abstrakt

Tato práce se zabývá variačními bayesovskými metodami. Získané poznatky byly následně aplikovány na směšové modely. Následně byly navrženy iterační algoritmy pro výpočet hledaných variačních parametrů, které byly posléze aplikovány na data. Kromě simulovaných dat byla k dispozici i data reálná.

Klíčová slova: variační bayes, bayesovská statistika, mean-field aproximace, směšové modely, ELBO

Abstract

This thesis is focused to variational bayesian methods. Informations gained about variational bayes were applied to mixture models. Then we implemented iterative algorithms used for computation of variational parameters. Algorithms were tested on simulated dataset and also on real dataset.

Key Words: variational bayes, bayesian statistics, mean-field approximation, mixture models, ELBO

Obsah

Seznam použitých zkratek a symbolů	8
Seznam obrázků	9
Seznam tabulek	10
1 Úvod	11
2 Základní pojmy	12
2.1 Teorie míry a pravděpodobnosti	12
2.2 Kullback–Leiblerova divergence	13
2.3 Exponenciální rodina	14
2.4 Směšové modely	16
3 Bayesovská statistika	18
3.1 Úvod k Bayesovské statistice	18
3.2 Odvození Bayesovského modelu	18
4 Variační bayesovské metody	22
4.1 Skryté lokální a globální parametry	22
4.2 <i>Mean-field</i> variační aproximace	23
4.3 Odhad parametrů směšového modelu se 2 komponentami	27
4.4 Odhad parametrů směšového modelu s n komponentami	36
5 Numerické experimenty	45
6 Závěr	65
Literatura	66
Přílohy	66
A Popis algoritmů	67

Seznam použitých zkratk a symbolů

$\mathcal{B}(\mathbb{R}^n)$	– borelovská σ -algebra (v \mathbb{R}^n)
\mathbb{R}^*	– rozšířená reálná čísla, $\mathbb{R} \cup \{+\infty, -\infty\}$
s.v.	– skoro všude
KL divergence	– Kullback-Leiblerova divergence
ELBO	– Evidence lower bound
MCMC	– Markov Chain Monte Carlo
\propto	– rovnost až na normalizační konstantu
$E[X]$	– střední hodnota náhodné veličiny X
$\mathcal{N}(\mu, \sigma^2)$	– normální (gaussovo) rozdělení s parametry μ, σ_1^2
$\mathcal{B}e(d_1, d_2)$	– beta rozdělení s parametry <i>shape</i> d_1 a <i>rate</i> d_2
$D(\alpha)$	– diskrétní rozdělení s parametrem $\alpha = (\alpha_1, \dots, \alpha_J)$, kde $\sum_{j=1}^J \alpha_j = 1$
$\mathcal{NG}(\mu, \tau, a, b)$	– normální-gamma rozdělení s parametry μ, τ, a, b
$\mathcal{D}ir(d)$	– Dirichletovo rozdělení s parametrem $d = d_1, \dots, d_J$

Seznam obrázků

1	Grafový model s pozorovanými daty x , lokálními skrytými veličinami c a globálními skrytými veličinami β	23
2	Algoritmus ELBO	27
3	Grafový směšový model se dvěma komponentami	27
4	Grafová reprezentace Bayesovského směšového modelu se 2 komponentami	28
5	Grafový směšový model s J komponentami	36
6	Grafová reprezentace Bayesovského směšového modelu s J komponentami	38
7	Histogram nasimulovaných dat	45
8	Histogram a hustota dat	47
9	Odhad parametru $\mu_{N_1}^*$	47
10	Odhad parametru $\mu_{N_2}^*$	48
11	Odhad parametru $\sigma_{N_1}^*$	48
12	Odhad parametru $\sigma_{N_2}^*$	49
13	Odhad parametru $d_{N_1}^*$	49
14	Odhad parametru $d_{N_2}^*$	50
15	Histogram nasimulovaných dat	51
16	Histogram a hustoty dat	53
17	Odhad parametru $\mu_{N_1}^*$	53
18	Odhad parametru $\mu_{N_2}^*$	54
19	Odhad parametru $\mu_{N_3}^*$	54
20	Odhad parametru $k_{N_1}^*$	55
21	Odhad parametru $k_{N_2}^*$	55
22	Odhad parametru $k_{N_3}^*$	56
23	Odhad parametru $d_{N_1}^*$	56
24	Odhad parametru $d_{N_2}^*$	57
25	Odhad parametru $d_{N_3}^*$	57
26	Odhad parametru a_N^*	58
27	Odhad parametru b_N^*	58
28	Histogram tlakových dat	59
29	Histogram a hustoty tlakových dat	60
30	Odhad parametru $\mu_{N_1}^*$	61
31	Odhad parametru $\mu_{N_2}^*$	61
32	Odhad parametru $k_{N_1}^*$	62
33	Odhad parametru $d_{N_1}^*$	62
34	Odhad parametru $d_{N_2}^*$	63
35	Odhad parametru a_N^*	63
36	Odhad parametru b_N^*	64

Seznam tabulek

1	Tabulka variačních parametrů	47
2	Tabulka variačních parametrů	52
3	Tabulka variačních parametrů tlakových dat	60

1 Úvod

Bayesovská statistika nachází široké využití v oblastech statistiky, ve kterých pracujeme s neurčitými znalostmi. Umožňuje aplikaci na reálné úlohy s velkým počtem dat.

V bayesovské statistice máme k dispozici apriorní informaci (předběžná informace, např. fyzikální model) o parametrech, která je vyjádřena pomocí apriorního rozdělení. K odhadu dat využíváme aposteriorní rozdělení, které v sobě nese apriorní informaci a informaci o datech, získanou jejich pozorováním. Výstupem je tedy kompletní informace o parametru včetně neurčitosti charakterizované rozdělením pravděpodobnosti. Spočítat aposteriorní rozdělení může být velmi náročné a proto je potřeba jej aproximovat.

Aproximaci aposteriorního rozdělení lze provést za použití Markov Chain Monte Carlo (MCMC) metod. Ty však bývají výpočetně náročné a proto využíváme jako alternativu variačních bayesovských metod, které poskytují analytické řešení aproximace aposteriorního rozdělení. Odvození aproximace je téměř vždy velmi pracné, jeho výsledkem jsou iterační vztahy. Samotný výpočet je však nesrovnatelně rychlejší než MCMC.

V první kapitole si objasníme některé pojmy z oblasti teorie pravděpodobnosti potřebné pro pochopení problematiky bayesovské statistiky. V kapitole druhé se seznámíme s bayesovskými metodami. Ve třetí kapitole se budeme zabývat variačními metodami, které aplikujeme na směšové modely. V kapitole poslední jsou obsaženy numerické experimenty. V příloze jsou popsány algoritmy sloužící k nalezení variačních parametrů daných směšových modelů. Algoritmus byl rovněž aplikován i na reálná data, konkrétně se jedná o krevní tlaky pacientů.

K výpočtům, simulaci dat a vykreslení obrázků byl použit statistický software R.

2 Základní pojmy

V této kapitole si objasníme některé pojmy z teorie pravděpodobnosti, které budou potřeba pro pochopení problematiky následujícího textu. Více lze nalézt například v [8], nebo v [7]. Je-li jasné z kontextu, značíme hustoty pravděpodobnosti bez indexů vyjadřujících příslušné náhodné veličiny. Ty budou jednoznačně určeny argumenty, např. $p_{X|Y}(x|y) \stackrel{\text{ozn.}}{=} p(x|y)$. Podobnou konvenci používáme i pro funkce, které jsou součástí hustot pravděpodobnosti.

Některé termíny v textu budou psány anglicky, jelikož není znám jejich český překlad. Takové termíny budou zapsány pomocí kurzívy.

2.1 Teorie míry a pravděpodobnosti

Bayesovská statistika je založena na teorii pravděpodobnosti a proto si definujeme některé pojmy. Jelikož budeme pracovat s hustotami pravděpodobnosti, potřebujeme definice absolutně spojitě míry, Radonovy-Nikodymovy derivace, indukované míry a další.

Definice 1 *Nechť (Ω, \mathcal{A}) je měřitelný prostor a μ_1, μ_2 jsou míry na (Ω, \mathcal{A}) . Nechť dále platí $\forall A \in \mathcal{A} : \mu_1(A) = 0 \Rightarrow \mu_2(A) = 0$.*

Pak říkáme, že míra μ_2 je absolutně spojitá vzhledem k míře μ_1 , značíme $\mu_2 \ll \mu_1$.

Míra μ_1 se nazývá dominující míra k míře μ_2 .

Věta 1 *Nechť (Ω, \mathcal{A}) je měřitelný prostor, μ_1, μ_2 jsou míry na \mathcal{A} , μ_1 je σ -konečná a $\mu_2 \ll \mu_1$.*

Pak existuje $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^, \mathcal{B}(\mathbb{R}^*))$ taková, že $\forall A \in \mathcal{A} : \mu_2(A) = \int_A f(x) d\mu_1(x)$.*

Funkce f je určena jednoznačně až na množinu μ_1 míry 0 (tj. $\forall A \in \mathcal{A} : \mu_2(A) = \int_A \tilde{f}(x) d\mu_1(x) \Rightarrow f = \tilde{f} \mu_1$ -s.v.). Je-li μ_2 σ -konečná, pak f je konečná μ_1 -s.v.

Je-li navíc $g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ integrovatelná, pak $\int g(x) d\mu_2(x) = \int g(x) f(x) d\mu_1(x)$.

Každou funkci f nazýváme Radonovou-Nikodymovou derivací μ_2 vzhledem k μ_1 , značíme $\frac{d\mu_2}{d\mu_1}$.

Symbolem \mathbb{R}^* značíme rozšířená reálná čísla, tj. $\mathbb{R} \cup \{+\infty, -\infty\}$. Jako $\mathcal{B}(\mathbb{R})$ se značí borelovská σ -algebra v \mathbb{R} .

Definice 2 *Nechť $(\Omega, \mathcal{A}, \mu)$ je prostor s mírou, (T, \mathcal{C}) je měřitelný prostor a $f : (\Omega, \mathcal{A}) \rightarrow (T, \mathcal{C})$. Definujeme $\nu(A) := \mu(f^{-1}(A))$ pro $A \in \mathcal{C}$. Míru ν na (T, \mathcal{C}) nazýváme mírou indukovanou zobrazením f z míry μ .*

$f : (\Omega, \mathcal{A}) \rightarrow (T, \mathcal{C})$ je zobrazení $f : \Omega \rightarrow T$, které je $\mathcal{A} - \mathcal{C}$ měřitelné.

Definice 3 *Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní prostor, $(\mathcal{X}, \mathcal{B}, \nu)$ je prostor s mírou a $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$.*

Předpokládejme dále, že pro indukovanou míru P_x platí $P_x \ll \nu$.

Pak každou Radonovu-Nikodymovu derivaci $\frac{dP_x}{d\nu}$ nazýváme hustotou pravděpodobnosti měřitelného zobrazení X vzhledem k míře ν .

Definice 4 Necht (Ω, \mathcal{A}, P) je pravděpodobnostní prostor. Zobrazení $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ se pak nazývá náhodná veličina. Necht náhodná veličina X je integrovatelná, tzn. $\int X dP$ existuje a je konečný. Pak $\int X dP$ nazveme střední hodnotou náhodné veličiny X , značíme $E(X)$.

Je-li $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, kde $X_1 \dots X_n$ jsou integrovatelné náhodné veličiny, pak střední hodnotou náhodného vektoru X rozumíme vektor $E(X) = (E(X_1), \dots, E(X_n))$.

Věta 2 (Lebesgueova) Necht $(\Omega, \mathcal{A}, \mu)$ je prostor s mírou, $\forall n \in \mathbb{N} : f_n : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$, $f_n \rightarrow f$ s.v., $\exists g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$ integrovatelná funkce taková, že $|f_n| \leq g$ s.v. Pak f je integrovatelná a $\int f_n d\mu \rightarrow \int f d\mu$.

2.2 Kullback–Leiblerova divergence

Kullback-Leiblerova divergence kvantifikuje odlišnost dvou pravděpodobnostních distribucí.

Definice 5 Mějme náhodnou veličinu X s hodnotami v $(\mathcal{X}, \mathcal{A})$. Necht P a Q jsou rozdělení pravděpodobnosti náhodné veličiny X , absolutně spojitě vzhledem k míře μ na $(\mathcal{X}, \mathcal{A})$. Označme hustoty $p = \frac{dP}{d\mu}$ a $q = \frac{dQ}{d\mu}$. Pak Kullback-Leiblerova divergence je definována jako

$$KL(p(x)||q(x)) = \int_{\mathcal{X}} \ln \frac{p(x)}{q(x)} p(x) d\mu(x) = E_p \left[\ln \left(\frac{p(x)}{q(x)} \right) \right]$$

pro $P \ll Q$.

Jestliže neplatí $P \ll Q$, pak $KL(p(x)||q(x)) = \infty$.

Vlastnosti Kullback-Leiblerovy divergence:

- obecně neplatí symetrie: $KL(p(x)||q(x)) = KL(q(x)||p(x))$
- neplatí trojúhelníková nerovnost:
 $KL(p(x)||r(x)) \geq KL(p(x)||q(x)) + KL(q(x)||r(x))$
- je nezáporná: $KL(p(x)||q(x)) \geq 0$
- $KL(p(x)||q(x)) = 0$ právě tehdy, když pro libovolné hustoty $p(x)$, $q(x)$ platí $p(x) = q(x)$, μ -s.v.
- je konvexní v obou argumentech: pro každé hustoty $p(x)$, $q(x)$, $r(x)$ a každé číslo $\lambda \in \langle 0, 1 \rangle$ platí:

$$KL(\lambda p(x) + (1 - \lambda) q(x)||r(x)) \leq \lambda KL(p(x)||r(x)) + (1 - \lambda) KL(q(x)||r(x))$$

$$KL(p(x)||\lambda q(x) + (1 - \lambda) r(x)) \leq \lambda KL(p(x)||q(x)) + (1 - \lambda) KL(p(x)||r(x))$$

- pro sdružené pravděpodobnostní distribuce p a q dvou spojitých náhodných veličin X a Y je divergence dvou sdružených distribucí rovna součtu divergence marginálních distribucí a střední hodnoty divergence podmíněných distribucí:

$$\begin{aligned}
KL(p(x, y) || q(x, y)) &= \int_X \int_Y p(x, y) \ln \frac{p(x, y)}{q(x, y)} d\mu(x, y) = \\
&= \int_X \int_Y p(x|y) p(y) \left(\ln \frac{p(x|y)}{q(x|y)} + \ln \frac{p(y)}{q(y)} \right) d\mu(x, y) = \\
&= KL(p(y) || q(y)) + E_{p(y)} [KL(p(x|y) || q(x|y))],
\end{aligned}$$

kde $KL(p(x|y) || q(x|y))$ je podmíněná KL-divergence.

Ačkoliv KL divergence měří "vzdálenost" distribucí, nesplňuje vlastnosti metriky.

Součet $KL(P||Q) + KL(Q||P)$ někdy nazýváme jako symetrickou Kullback-Leiblerovou divergenci.

2.3 Exponenciální rodina

Exponenciální rodina je množina pravděpodobnostních rozdělení ve specifickém tvaru, viz. definice 6 .

Definice 6 *Uvažujme parametrickou rodinu hustot $p_X(x|\theta)$ vzhledem k σ -konečné míře μ na $(\mathcal{X}, \mathcal{A})$. Necht' pro nějaké $n \in \mathbb{N}$ máme vektor parametrů $\theta = (\theta_1, \dots, \theta_n)$ z parametrického prostoru Π a necht' dále existují funkce $t_1, \dots, t_n : \mathcal{X} \rightarrow \mathbb{R}$, $\eta_1, \dots, \eta_n : \Pi \rightarrow \mathbb{R}$, $a : \Pi \rightarrow \mathbb{R}$, $h : \mathcal{X} \rightarrow \mathbb{R}_+$, takové, že platí*

$$p_X(x|\theta) = h(x) \exp \left(\sum_{i=1}^n \eta_i^T(\theta) t_i(x) - a(\theta) \right). \quad (1)$$

Pak parametrickou rodinu s hustotou pravděpodobnosti $p_X(x|\theta)$ nazýváme exponenciální rodinou. Funkci $t(x) = (t_1(x), \dots, t_n(x))$ nazýváme postačující statistikou rozdělení.

Předchozí vztah zkráceně zapíšeme jako

$$p_X(x|\theta) = h(x) \exp \left(\eta^T(\theta) t(x) - a(\theta) \right). \quad (2)$$

Jelikož integrál hustoty pravděpodobnosti přes celý obor hodnot náhodné veličiny X musí být roven jedné dostáváme integrací rovnice (2) vztah.

$$\int_{\mathcal{X}} h(x) \exp \left(\eta^T(\theta) t(x) - a(\theta) \right) d\mu = 1.$$

Tímto je jednoznačně určena log normalizační funkce $a(\theta)$:

$$a(\theta) = \ln \left(\int_{\mathcal{X}} h(x) \exp \left(\eta^T(\theta) t(x) \right) d\mu \right)$$

Definice 7 Uvažujme exponenciální rodinu (2) z definice 6. Parametrizaci

$$p_X(x|\lambda) = h(x) \exp(\lambda^T t(x) - a(\lambda))$$

nazýváme přirozenou parametrizací exponenciální rodiny. Parametr λ nazýváme přirozeným parametrem a množinu

$$\Lambda = \left\{ \lambda : \left(\int_{\mathcal{X}} h(x) \exp(\lambda^T t(x)) d\mu \right) < \infty \right\}$$

nazýváme množinou přirozených parametrů exponenciální rodiny $p_X(x|\lambda)$.

Pravděpodobnostní rozdělení mohou být reprezentována mnoha různými parametrizacemi. Uvažujme bijektivní zobrazení $\pi : \Lambda \rightarrow \tilde{\Lambda}$ a označme $\eta(\theta) = \pi^{-1}(\theta)$ pro $\theta \in \tilde{\Lambda}$. Pak exponenciální rodinu $p_X(x|\lambda)$ lze ekvivalentně vyjádřit ve tvaru

$$p_X(x|\theta) = p(x|\lambda) |_{\lambda=\eta(\theta)}.$$

Gradient $a(\lambda)$

Derivací funkce $a(\lambda)$ dostáváme střední hodnotu postačující statistiky:

$$\begin{aligned} \nabla_{\lambda} a(\lambda) &= \nabla_{\lambda} \left(\ln \int \exp(\lambda t(x)) h(x) d\mu \right) = \\ &= \frac{\nabla_{\lambda} \int \exp(\lambda t(x)) h(x) d\mu}{\int \exp(\lambda t(x)) h(x) d\mu} = \\ &= \frac{\int t(x) \exp(\lambda t(x)) h(x) d\mu}{\int \exp(\lambda t(x)) h(x) d\mu} = \\ &= E_{\lambda}[t(X)], \end{aligned}$$

kde E_{λ} značí střední hodnotu vzhledem k $p(x|\eta)$. Gradient ∇_{λ} lze přesunout dovnitř integrálu, díky Lebesgueově větě.

Příklad 1

Hustotu normálního rozdělení lze zapsat v exponenciálním tvaru:

$$\begin{aligned} p(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln \sigma\right) = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^T (x, x^2) - \left(\frac{\mu^2}{2\sigma^2} + \ln \sigma\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left((\lambda_1, \lambda_2)^T (x, x^2) - \left(-\frac{\lambda_1^2}{4\lambda_2} - \frac{1}{2} \ln(-2\lambda_2)\right)\right) \end{aligned}$$

s parametry:

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \lambda = \begin{pmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{pmatrix},$$

postačující statistikou

$$t(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix},$$

log normalizační funkcí

$$a(\lambda) = -\frac{\lambda_1^2}{4\lambda_2} - \frac{1}{2} \ln(-2\lambda_2),$$

a funkcí

$$h(\lambda) = \frac{1}{\sqrt{2\pi}}.$$

Vztah mezi běžnou a přirozenou parametrizací je:

$$\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} -\frac{\lambda_1}{2\lambda_2} \\ -\frac{1}{2\lambda_2} \end{pmatrix}$$

Log normalizační funkce pak má tvar

$$a(\mu, \sigma^2) = -\frac{\mu^2}{2\sigma^2} + \ln \sigma \quad (3)$$

■

2.4 Směšové modely

Definice 8 *Hustota p náhodného vektoru X se nazývá pravědpodobnostní směs s C komponentami jestliže*

$$p(x|\theta) = \sum_{c=1}^C \alpha_c m_c(x|\theta_c),$$

kde m_c jsou hustoty komponent směsi pocházející ze stejné parametrické rodiny s lišícími se parametry a α_c jsou váhy směsi, pro které platí: $\alpha_c \geq 0$, $\sum_c \alpha_c = 1$. Dále θ_c je parametr, nebo vektor parametrů c -té komponenty a θ obsahuje všechny parametry směšového modelu:

$$\theta = (\alpha_1, \dots, \alpha_C, \theta_1, \dots, \theta_C).$$

Směšový model také můžeme chápat jako marginální hustotu $p(x|\theta)$ sdružené hustoty

$$p(x, c|\theta) = p(c|\theta) p(x|c, \theta),$$

kde $p(c|\theta) = \alpha_c$ a $p(x|c, \theta) = m_c(x|\theta_c)$.

Jako příklad uvedeme N –rozměrnou gaussovu směs, která je dána vztahem

$$p(x|\theta) = \sum_{c=1}^C \alpha_c m_c(x|\mu_c, \Sigma_c),$$

kde $m_c(x|\mu_c, \Sigma_c)$ jsou hustoty N –rozměrného gaussova rozdělení.

$$m_c(x|\mu_c, \Sigma_c) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_c|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\right).$$

$\mu_c \in \mathbb{R}^N$ je vektor středních hodnot a $\Sigma_c \in \mathbb{R}^{N,N}$ jsou kovarianční matice. V případě, že dimenze $N = 2$ je gaussova směs definována jako

$$p(x|\theta) = \alpha_1 m_1(x|\mu_1, \sigma_1^2) + (1 - \alpha_1) m_2(x|\mu_2, \sigma_2^2)$$

se středními hodnotami μ_1, μ_2 a rozptyly σ_1^2, σ_2^2 .

3 Bayesovská statistika

3.1 Úvod k Bayesovské statistice

Ve statistice jsou používány dva typy přístupů - klasická a Bayesovská statistika. Rozdíl mezi nimi je v odlišné interpretaci pojmu pravděpodobnosti a tedy odlišných způsobů vyvozování závěrů o dané populaci. Dále reprezentace neznámých parametrů a potřeba znalosti předběžné informace o parametrech.

Parametry pravděpodobnostních rozdělení jsou v klasické statistice známe nebo neznámé konstanty. U Bayesovského přístupu jsou parametry chápány jako náhodné veličiny s pravděpodobnostním rozdělením, které popisuje znalost o hodnotách neznámých parametrů. Pravděpodobnostní rozdělení těchto parametrů vyjadřuje nejistotu o jejich hodnotě. Při odhadu parametrů využíváme jejich předběžné znalosti společně s informací obsaženou v datech viz. například [6].

3.2 Odvození Bayesovského modelu

Mějme náhodnou veličinu X s neznámým parametrem θ . Sdruženou hustotu $p(x, \theta)$ lze z definice zapsat jako $p(x, \theta) = p(x|\theta)p(\theta)$.

Při pevném x , jako funkce parametru θ , se funkce $p(x|\theta)$ nazývá věrohodnostní funkce a popisuje náš statistický model. Poskytuje informace o parametru, která je obsažená v datech.

Apriorní hustota $p(\theta)$ vyjadřuje znalosti o parametru ještě před tím, než byla získána data. Tyto informace mohou být získány například z podobných předchozích experimentů nebo odpovídajícího fyzikálního modelu. Jelikož apriorní rozdělení významně ovlivňuje výsledné rozdělení, je jeho volba důležitá. Nalézt vhodné apriorní rozdělení je však někdy velmi obtížné.

Aposteriorní hustota v sobě nese jak informace o parametrech z dat, tak apriorní informaci. Poskytuje nám tedy souhrnou informaci o neznámém parametru θ . Tato hustota je definována jako

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}. \quad (4)$$

Aposteriorní rozdělení využíváme ke zkoumání skryté struktury dat, abychom mohli např. odhadnout, jak budou vypadat následující data. Marginální hustotu $p(x)$ dostaneme integrací sdružené hustoty $p(x, \theta)$

$$p(x) = \int p(x, \theta) d\theta = \int p(x|\theta)p(\theta) d\theta.$$

Funkce $p(x)$ ve vztahu (4) hraje roli normalizační konstanty, pro kterou platí $\int p(x|\theta)p(\theta) d\theta = p(x)$. Rovnici (4) lze tedy zapsat ve tvaru

$$p(\theta|x) \propto p(x|\theta)p(\theta),$$

kde \propto vyjadřuje rovnost až na multiplikativní konstantu.

Aposteriorní hustota je tedy dána součinem věrohodnostní funkce a apriorní hustoty. V případě, že aposteriorní rozdělení patří do stejné třídy rozdělení jako apriorní rozdělení, pak jsou tato rozdělení nazývána jako konjugována rozdělení k danému modelu. Třída těchto rozdělení se nazývá konjugovaný systém rozdělení.

Příklad 2

Nyní ukážeme odvození aposteriorního rozdělení parametru μ , přičemž rozptyl σ^2 je znám.

Nechť $x = (x_1, \dots, x_N)$ jsou data, tj. hodnoty nezávislých stejně rozdělených náhodných veličin s rozdělením $\mathcal{N}(\mu, \sigma^2)$, kde parametr μ není znám. Pak věrohodnostní funkce je:

$$p(x|\mu, \sigma^2) = \prod_{n=1}^N p(x_n|\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2\right). \quad (5)$$

Definujeme výběrový průměr a rozptyl

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$$

$$s^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2$$

Výraz v exponenciále v rovnici (5) upravíme takto

$$\begin{aligned} \sum_{n=1}^N (x_n - \mu)^2 &= \sum_{n=1}^N (x_n - \bar{x} - \mu + \bar{x})^2 = \\ &= \sum_{n=1}^N (x_n - \bar{x})^2 + \sum_{n=1}^N (\bar{x} - \mu)^2 - 2 \sum_{n=1}^N (x_n - \bar{x})(\mu - \bar{x}) = \\ &= Ns^2 + N(\bar{x} - \mu)^2. \end{aligned}$$

Poslední rovnost plyne z

$$\begin{aligned} \sum_{n=1}^N (x_n - \bar{x})(\mu - \bar{x}) &= (\mu - \bar{x}) \left(\sum_{n=1}^N x_n - N\bar{x} \right) = \\ &= (\mu - \bar{x})(N\bar{x} - N\bar{x}) = \\ &= 0. \end{aligned}$$

Z předchozích vztahů vyplývá

$$p(x|\mu, \sigma^2) = (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} (Ns^2 + N(\bar{x} - \mu)^2)\right) \propto$$

$$\propto (\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{Ns^2}{2\sigma^2}\right) \exp\left(-\frac{N(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

V případě známého rozptylu σ^2 je model náhodné veličiny X určen podmíněnou hustotou

$$p(x|\mu) \propto \exp\left(-\frac{N(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

Apriorní rozdělení zvolíme jako normální rozdělení

$$p(\mu) \propto \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right), \text{ tj.}$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

Apsteriorní rozdělení

Pro apsteriorní rozdělení platí:

$$p(\mu|x) \propto p(x|\mu) p(\mu) \propto$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2\right) \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right) =$$

$$= \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n^2 + \mu^2 - 2x_n\mu) - \frac{1}{2\sigma_0^2} (\mu^2 + \mu_0^2 - 2\mu_0\mu)\right).$$

$$p(\mu|x) \propto \exp\left(-\frac{\mu^2}{2} \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right) + \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{n=1}^N x_n}{\sigma^2}\right) - \left(\frac{\mu_0^2}{2\sigma_0^2} + \frac{\sum_{n=1}^N x_n^2}{2\sigma^2}\right)\right)$$

Z předchozího zápisu lze vidět, že se jedná o hustotu normálního rozdělení. Označme si střední hodnoty a rozptyly tohoto rozdělení jako μ_n , a σ_n^2 . Pak

$$p(\mu|x) \propto \exp\left(-\frac{1}{2\sigma_n^2} (\mu^2 - 2\mu\mu_n + \mu_n^2)\right) =$$

$$= \exp\left(-\frac{1}{2\sigma_n^2} (\mu - \mu_n)^2\right).$$

Porovnáním koeficientů získáme nový rozptyl σ_n^2

$$\begin{aligned} -\frac{\mu^2}{2\sigma_n^2} &= -\frac{\mu^2}{2} \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \right) \\ \frac{1}{\sigma_n^2} &= \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \\ \sigma_n^2 &= \frac{\sigma^2 \sigma_0^2}{N\sigma_0^2 + \sigma^2} = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}. \end{aligned}$$

Porovnáním μ dostaneme střední hodnotu aposteriorního rozdělení μ_n

$$\begin{aligned} \frac{-2\mu\mu_n}{-2\sigma_n^2} &= \mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{n=1}^N x_n}{\sigma^2} \right) \\ \frac{\mu_n}{\sigma_n^2} &= \frac{\sum_{n=1}^N x_n}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} = \\ &= \frac{\sigma_0^2 N \bar{x} + \sigma^2 \mu_0}{\sigma^2 \sigma_0^2}. \end{aligned}$$

A proto

$$\mu_n = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} \bar{x} = \sigma_n^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{N\bar{x}}{\sigma^2} \right).$$

Aposteriorní rozdělení je tedy normální. Střední hodnota má proto normální rozdělení:

$$\mu \sim \mathcal{N}(\mu_n, \sigma_n^2).$$

■

4 Variační bayesovské metody

Jedním z problému Bayesovské statistiky je výpočet aposteriorní hustoty pravděpodobnosti, který může být mnohdy obtížný.

K aproximaci aposteriorního rozdělení jsou využívány Markov chain Monte Carlo metody a Variační metody. V práci se budeme dále zabývat pouze variačními metodami.

Budeme uvažovat pravděpodobnostní model skrytých náhodných veličin (parametrů) a pozorovaných náhodných veličin (dat) popsáný grafovým modelem, viz. [5], pro něj počítáme aposteriorní rozdělení. Vztah mezi veličinami je popsán modelem a apriorním rozdělením. Cílem variačních bayesovských metod je odvození aproximace aposteriorního rozdělení.

Variační aproximace spočívá ve volbě vhodné třídy rozdělení pro skryté veličiny. Z této třídy pak hledáme v nějakém smyslu nejbližší rozdělení k aposteriornímu rozdělení. Na tento problém lze tedy pohlížet jako na optimalizační úlohu.

4.1 Skryté lokální a globální parametry

Uvažujme grafový model s konkrétní nezávislostní strukturou z [3]¹ s N náhodnými veličinami $x = (x_1, \dots, x_N)$, s N lokálními skrytými náhodnými veličinami $c = (c_1, \dots, c_N)$, kde každé c_n je vektor $c_n = (c_{n,1}, \dots, c_{n,J})$, vektorem globálních parametrů β a vektorem pevně daných parametrů α . Situace je popsána grafovým modelem na obrázku 1, kterým popisujeme podmíněné nezávislosti mezi proměnnými, více viz. [5].

Sdruženou pravděpodobnost tohoto modelu lze vyjádřit vztahem:

$$p(x, c, \beta | \alpha) = p(\beta | \alpha) \prod_{n=1}^N p(x_n, c_n | \beta). \quad (6)$$

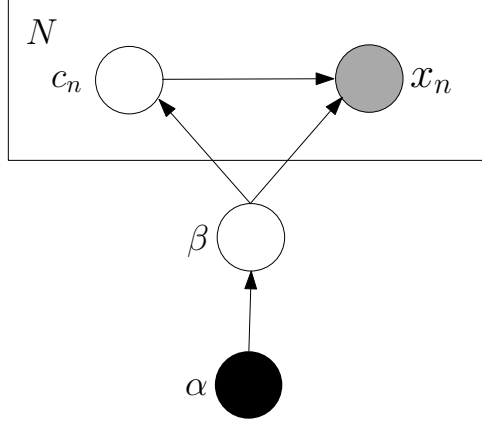
Rozdíl mezi lokálními a globálními skrytými veličinami je dán podmíněnými nezávislostmi. V uvažovaném modelu je vektor (x_n, c_n) podmíněně nezávislý na vektoru (x_{-n}, c_{-n}) , je-li znám vektor globálních parametrů β a pevně daných parametrů α , tj.

$$p(x_n, c_n | x_{-n}, c_{-n}, \beta, \alpha) = p(x_n, c_n | \beta, \alpha),$$

kde x_{-n} a c_{-n} jsou náhodné vektory, jejichž složkami jsou všechny složky x, c kromě n -té. Globální parametry β mají své apriorní rozdělení $p(\beta | \alpha)$.

Plně podmíněné rozdělení je definováno jako rozdělení skryté veličiny podmíněné dalšími skrytými veličinami a pozorováními. Budeme předpokládat, že tato rozdělení patří do exponenciální

¹řada modelů používaných v praxi má právě tuto strukturu (např. směsi)



Obrázek 1: Grafový model s pozorovanými daty x , lokálními skrytými veličinami c a globálními skrytými veličinami β

rodiny:

$$p(\beta|x, c, \alpha) = h_g(\beta) \exp\left(\eta_g(x, c, \alpha)^T t_g(\beta) - a_g(\eta_g(x, c, \alpha))\right) \quad (7)$$

$$p(c_{nj}|x_n, c_{n,-j}, \beta) = h_l(c_{nj}) \exp\left(\eta_l(x_n, c_{n,-j}, \beta)^T t_l(c_{nj}) - a_l(\eta_l(x_n, c_{n,-j}, \beta))\right). \quad (8)$$

Dostáváme tak tvar rozdělení (x_n, c_n)

$$p(x_n, c_n|\beta) = h(x_n, c_n) \exp\left(\beta^T t(x_n, c_n) - a(\beta)\right)$$

a apriorního rozdělení β

$$p(\beta) = h(\beta) \exp\left(\alpha^T t(\beta) - a(\alpha)\right),$$

kde $t(\beta) = (\beta, -a(\beta))$ Naším cílem bude aproximovat aposteriorní rozdělení skrytých veličin:

$$p(c, \beta|x) = \frac{p(x, c, \beta)}{\int p(x, c, \beta) dc d\beta}.$$

4.2 Mean-field variační aproximace

Cílem variačního odvození je aproximovat hustotu skrytých veličin podmíněnou pozorovanými veličinami rozdělením ze zvolené třídy. Na tento problém pohlížíme jako na optimalizační úlohu. Zavedeme rodinu rozdělení skrytých veličin parametrizovanou variačními parametry. Abychom našli prvek rodiny, který je nejbližší aposteriornímu rozdělení musíme tyto parametry optimalizovat, tak, aby byl tento prvek co nejbližší aposteriornímu rozdělení. Vzdálenost těchto rozdělení měříme KL divergencí. Výsledné rozdělení, nazývané variační rozdělení, používáme jako aproximaci aposteriorního rozdělení. Toto aproximované aposteriorní rozdělení označme $q(\beta, c)$.

V případě *mean-field* variační aproximace budeme požadovat vzájemnou nezávislost skrytých veličin, tj. $q(\beta, c) = q(\beta) \prod_{n=1}^N q(c_n)$.

The evidence lower bound

Při variačním odvozování minimalizujeme KL divergenci variačního rozdělení a posteriorního rozdělení. Problém hledání tohoto maxima je ekvivalentní k nalezení prvku rodiny nejbližšího ve smyslu KL divergence posteriornímu rozdělení:

$$\begin{aligned} KL(q(c, \beta) || p(c, \beta | x)) &= E_q [\ln q(c, \beta)] - E_q [\ln p(c, \beta | x)] \\ &= E_q [\ln q(c, \beta)] - E_q [\ln p(x, c, \beta)] + \ln p(x) \\ &= -\mathcal{L}(q) + k, \end{aligned} \tag{9}$$

kde $\ln p(x)$ je nahrazeno konstantou k , protože nezávisí na q . Záporná funkce $-\mathcal{L}(q)$ je označována jako *The evidence lower bound* (ELBO).

ELBO obsahuje dva výrazy. První je střední hodnota logaritmu sdružené pravděpodobnosti $E_q [\ln p(x, c, \beta)]$ a druhý je entropie variačního rozdělení $-E_q [\ln q(c, \beta)]$. Oba tyto výrazy jsou závislé na variačním rozdělení skrytých veličin $q(c, \beta)$.

Budeme požadovat, aby $q(c, \beta)$ pocházelo z takové rodiny, pro kterou budeme schopni efektivně spočítat střední hodnoty ELBO. Následně najdeme takový prvek rodiny, který maximalizuje ELBO. Optimální rozdělení pak použijeme k aproximaci posteriorního.

ELBO lze také interpretovat jako dolní hranici logaritmu marginální pravděpodobnosti pozorovaných dat $\ln p(x)$. Odvodíme jej za pomoci rozdělení skrytých veličin $q(c, \beta)$ a Jensenovy nerovnosti, viz. [7]. Tímto dostáváme následující hranici logaritmu marginální pravděpodobnosti:

$$\begin{aligned} \ln p(x) &= \ln \int p(x, c, \beta) dcd\beta \\ &= \ln \int p(x, c, \beta) \frac{q(c, \beta)}{q(c, \beta)} dcd\beta \\ &= \ln \left(E_q \left[\frac{p(x, c, \beta)}{q(c, \beta)} \right] \right) \\ &\geq E_q [\ln p(x, c, \beta)] - E_q [\ln q(c, \beta)] \\ &\stackrel{\text{def.}}{=} \mathcal{L}(q) \end{aligned}$$

Z Jensenovy nerovnosti a konkávnosti logaritmické funkce plyne pro každou integrovatelnou náhodnou veličinu Y platnost nerovnosti $\ln E[Y] \geq E[\ln(Y)]$.

Druhý výraz ve vztahu (9) je entropie globálního variačního rozdělení. První výraz je střední hodnotou logaritmu sdružené věrohodnostní funkce, který můžeme upravit tak, abychom oddělili výrazy nezávislé na β od výrazů na nich závislých:

$$E_q [\ln p(x, c, \beta)] = E_q [\ln p(x, c)] + E_q [\ln p(\beta | x, c)].$$

Výraz $E_q [\ln p(x, c)]$ nezávisí na β a bude tedy zahrnut do konstanty k

$$E_q [\ln p(x, c, \beta)] = E_q [\ln p(\beta|x, c)] + k.$$

Mean-field variační rodina

Nejjednodušší variační rodinou rozdělení je *mean-field* rodina, ve které jsou skryté veličiny vzájemně nezávislé a jsou určeny svými parametry:

$$q(c, \beta) = q(\beta|\lambda) \prod_{n=1}^N \prod_{j=1}^J q(c_{nj}|\phi_{nj}). \quad (10)$$

Globální parametry λ určují globální veličiny a lokální parametry ϕ_n určují lokální veličiny v n -té složce. ELBO je funkcí těchto parametrů.

Předchozí rovnice určuje tvar variační rodiny, ale nespecifikuje její formu. Budeme požadovat, aby byly $q(\beta|\lambda)$ a $q(c_{nj}|\phi_{nj})$ ve stejné exponenciální rodině jako jsou plně podmíněná rozdělení $p(\beta|x, c)$ a $p(c_{nj}|x_n, c_{n,-j}, \beta)$ v rovnici (7) a (8). Variační parametry λ a ϕ_{nj} jsou přirozenými parametry těchto rodin:

$$q(\beta|\lambda) = h(\beta) \exp(\lambda^T t_g(\beta) - a_g(\lambda)) \quad (11)$$

$$q(c_{nj}|\phi_{nj}) = h(c_n) \exp(\phi_{nj}^T t_l(c_{nj}) - a_l(\phi_{nj})). \quad (12)$$

Z předpokladu, že tyto exponenciální rodiny odpovídají exponenciálním rodinám plně podmíněným rozdělení vyplývá, že funkce t a h v rovnici (11) jsou stejné jako funkce t a h v rovnici (7). A obdobně t a h z rovnice (12) jsou stejné jako v rovnici (8), detaily viz. [3]. V následujícím textu budeme značit $q(c_{nj}|\phi_{nj})$ jako $q(c_{nj})$ a $q(\beta|\lambda)$ jako $q(\beta)$.

Gradient ELBO

V rovnici (9) jsme si definovali cílovou funkci \mathcal{L} a v rovnicích (10), (11) a (12) variační rodiny. Naším cílem je optimalizace cílové funkce \mathcal{L} vzhledem k variačním parametrům, čehož dosáhneme minimalizací funkce $\mathcal{L}(\lambda, \phi)$. Pomocí algoritmu najdeme lokální minima funkce $\mathcal{L}(\lambda, \phi)$, což jsou hledané variační parametry, více viz. [9]. Počáteční aproximaci variačních parametrů volíme náhodně. V jedné iteraci algoritmu postupně optimalizujeme každý variační parametr, ostatní parametry zůstanou zafixovány.

Nejdříve aktualizujeme parametr λ vzhledem k variačnímu rozdělení globálních veličin $q(\beta|\lambda)$. \mathcal{L} zapíšeme jako funkci λ , parametr ϕ je zafixován:

$$\mathcal{L}(\lambda) = E_q [\ln p(\beta|x, c)] - E_q [\ln q(\beta)] + k. \quad (13)$$

První dva výrazy jsou střední hodnoty zahrnující parametr β . Konstanta k obsahuje veličiny, které nezávisí na λ , ale pouze na rozdělení ostatních skrytých veličin. Tyto veličiny tedy nezávisí ani na $q(\beta|\lambda)$, protože ve variační rodině jsou všechny veličiny nezávislé.

Nyní provedeme substituci výrazu $q(\beta|\lambda)$ z rovnice (11) do rovnice (13), čímž získáme konečný výraz ELBO jako funkce proměnné λ :

$$\mathcal{L}(\lambda) = E_q[\eta_g(x, c, \alpha)]^T \nabla_\lambda a_g(\lambda) - \lambda^T \nabla_\lambda a_g(\lambda) + a_g(\lambda) + k. \quad (14)$$

Na pravé straně rovnice jsme využili toho, že střední hodnota postačující statistiky je rovna gradientu log normalizační funkce: $E_q[t_g(\beta)] = \nabla_\lambda a_g(\lambda)$. V konstantě k je obsažena střední hodnota $E_q[a_g(\eta_g(x, c, \alpha))]$, která nezávisí na $q(\beta)$.

Aktualizaci parametru λ získáme z gradientu

$$\nabla_\lambda \mathcal{L} = \nabla_{\eta_g}^2 a_g(\lambda) (E_q[\eta_g(x, c, \alpha)] - \lambda).$$

Tento gradient bude roven nule, položíme-li

$$\lambda = E_q[\eta_g(x, c, \alpha)]. \quad (15)$$

Nové globální variační parametry jsou tedy rovny střední hodnotě přirozeného parametru plně podmíněného rozdělení.

Obdobně vyjádříme parametry ϕ_{nj} z funkce

$$\mathcal{L}(\phi_{nj}) = E_q[\eta_l(x_n, c_{n,-j}, \beta)]^T \nabla_{\phi_{nj}} a_l(\phi_{nj}) - \phi_{nj}^T \nabla_{\phi_{nj}} a_l(\phi_{nj}) + a_l(\phi_{nj}) + k. \quad (16)$$

Ty dostaneme za pomoci gradientu:

$$\nabla_{\phi_{nj}} \mathcal{L} = \nabla_{\phi_{nj}}^2 a_l(\phi_{nj}) (E_q(\eta_l(x, c_{n,-j}, \beta)) - \phi_{nj}),$$

který je roven nule pro

$$\phi_{nj} = E_q[\eta_l(x_n, c_{n,-j}, \beta)] \quad (17)$$

Na základě rovnic (15) a (17) sestavíme algoritmus, kde iterujeme všechny lokální a globální parametry. Pro zastavení využijeme vhodného zastavovacího kritéria, např. pomocí prahu pro vzdálenost dvou po sobě jdoucích aproximací. S pomocí tohoto algoritmu nalezneme optimální parametry. Algoritmus je popsán na obrázku 2.

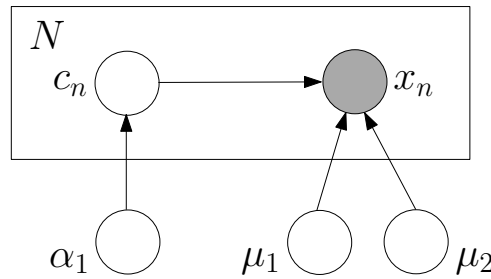
```

1: inicializace počátečního  $\lambda^{[1]}$ 
2: while dokud není splněna zastavovací podmínka do
3:   for pro  $n = 1 : N$  do
4:     aktualizace parametru  $\phi_{n,j}^{[i]} = E_{q^{[i-1]}} [\eta_l (x_n, c_n, -j, \beta)]$ 
5:   end for
6:   aktualizace parametru  $\lambda^{[i]} = E_{q^{[i]}} [\eta_g (x, c, \alpha)]$ 
7: end while

```

Obrázek 2: Algoritmus ELBO

4.3 Odhad parametrů směšového modelu se 2 komponentami



Obrázek 3: Grafový směšový model se dvěma komponentami

Nyní uvažujme směšový model se 2 gaussovskými komponentami se známými rozptyly σ_1^2, σ_2^2 . V tomto případě jsou skrytými veličinami střední hodnoty μ_1, μ_2 , váha 1. komponenty α_1 (váha 2. komponenty $1 - \alpha_1$) a čísla komponent c_n , kde c_n nabývají hodnot 1 nebo 2. V případě, že $c_n = 1$, pocházejí data z první komponenty, v případě že $c_n = 2$ pocházejí data z komponenty druhé. Grafový model pro tuto směs je znázorněn na obrázku 3.

Data x_n pocházejí z normálního rozdělení:

$$\begin{aligned}
x_n | \mu_1, \mu_2, \alpha_1, c_n &\sim \mathcal{N}(x_n | \mu_{c_n}, \sigma_{c_n}^2) = \frac{1}{\sqrt{2\pi\sigma_{c_n}^2}} \exp\left(-\frac{(x_n - \mu_{c_n})^2}{2\sigma_{c_n}^2}\right) = \\
&= \prod_{j=1}^2 \left(\frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_n - \mu_j)^2}{2\sigma_j^2}\right) \right)^{\delta(j, c_n)}
\end{aligned}$$

V následujícím textu budeme někdy značit hustotu pravděpodobnostní veličiny x s normálním rozdělením $\mathcal{N}(\mu, \sigma^2)$ jako $\mathcal{N}(x | \mu, \sigma^2)$. S využitím podmíněných nezávislostí dostáváme

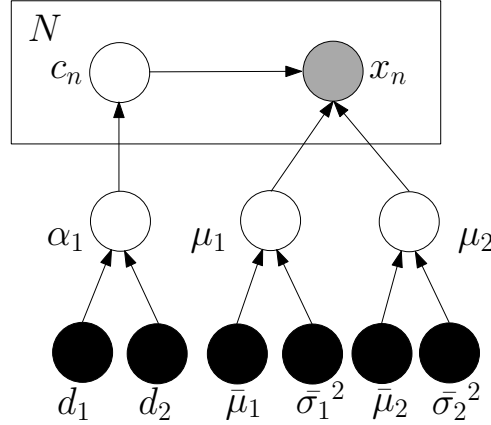
sdruženou pravděpodobnost, která je dána vztahem

$$\begin{aligned}
p(x, c, \mu_1, \mu_2, \alpha_1) &= p(\mu_1, \mu_2, \alpha_1) \prod_{n=1}^N p(x_n, c_n | \mu_1, \mu_2, \alpha_1) = \\
&= p(\mu_1, \mu_2, \alpha_1) \prod_{n=1}^N p(c_n | \alpha_1) p(x_n | \mu_1, \mu_2, c_n) = \\
&= p(\mu_1, \mu_2, \alpha_1) \prod_{n=1}^N \alpha_{c_n} \mathcal{N}(x_n | \mu_{c_n}, \sigma_{c_n}^2) \\
&= p(\mu_1, \mu_2, \alpha_1) \prod_{n=1}^N \left(\alpha_1 \mathcal{N}(x_n | \mu_1, \sigma_1^2) \right)^{\delta(c_n, 1)} \left((1 - \alpha_1) \mathcal{N}(x_n | \mu_2, \sigma_2^2) \right)^{\delta(c_n, 2)}
\end{aligned}$$

Váha α_1 pochází z beta rozdělení, střední hodnoty μ_1, μ_2 z normálního a hodnoty c_n z diskrétního rozdělení:

$$\begin{aligned}
\mu_1 &\sim \mathcal{N}(\bar{\mu}_1, \bar{\sigma}_1^2), p(\mu_1 | \bar{\mu}_1, \bar{\sigma}_1^2) = \frac{1}{\sqrt{2\pi\bar{\sigma}_1^2}} \exp\left(-\frac{(\mu_1 - \bar{\mu}_1)^2}{2\bar{\sigma}_1^2}\right) \\
\mu_2 &\sim \mathcal{N}(\bar{\mu}_2, \bar{\sigma}_2^2), p(\mu_2 | \bar{\mu}_2, \bar{\sigma}_2^2) = \frac{1}{\sqrt{2\pi\bar{\sigma}_2^2}} \exp\left(-\frac{(\mu_2 - \bar{\mu}_2)^2}{2\bar{\sigma}_2^2}\right) \\
\alpha_1 &\sim \mathcal{Be}(d_1, d_2), p(\alpha_1 | d_1, d_2) = \frac{(1 - \alpha_1)^{d_2-1} \alpha_1^{d_1-1}}{B(d_1, d_2)} \\
c_n &\sim D(\alpha_1), p(c_n | \alpha_1) = \alpha_{c_n} = \alpha_1^{\delta(c_n, 1)} (1 - \alpha_1)^{\delta(c_n, 2)}
\end{aligned}$$

Grafovou reprezentaci pro tento Bayesovský model najdeme na obrázku 4.



Obrázek 4: Grafová reprezentace Bayesovského směšového modelu se 2 komponentami

Nyní si vyjádříme plně podmíněné rozdělení globálních parametrů μ_1, μ_2, α_1 :

$$\begin{aligned}
p(\mu_1, \mu_2, \alpha_1 | x, c) &= \frac{p(\mu_1, \mu_2, \alpha_1, x, c)}{\int p(\mu_1, \mu_2, \alpha_1, x, c) d\mu_1 d\mu_2 d\alpha_1} \propto \\
&\propto p(\mu_1) p(\mu_2) p(\alpha_1) \prod_{n=1}^N p(c_n | \alpha_1) p(x_n | \mu_1, \mu_2, \alpha_1, c_n) = \\
&= p(\mu_1) p(\mu_2) p(\alpha_1) \prod_{n=1}^N p(c_n | \alpha_1) p(x_n | \mu_1, \mu_2, c_n) \\
p(\alpha_1) \prod_{n=1}^N p(c_n | \alpha_1) &= \frac{(1 - \alpha_1)^{d_2 - 1} \alpha_1^{d_1 - 1}}{B(d_1, d_2)} \alpha_1^{\sum_{n=1}^N \delta(c_n, 1)} (1 - \alpha_1)^{\sum_{n=1}^N \delta(c_n, 2)} = \\
&= \frac{(1 - \alpha_1)^{d_2 + \sum_{n=1}^N \delta(c_n, 2) - 1} \alpha_1^{d_1 + \sum_{n=1}^N \delta(c_n, 1) - 1}}{B(d_1, d_2)} \propto \\
&\propto (1 - \alpha_1)^{d_2 + \sum_{n=1}^N \delta(c_n, 2)} \alpha_1^{d_1 + \sum_{n=1}^N \delta(c_n, 1)},
\end{aligned}$$

tj.

$$\alpha_1 | x, c, \mu_1, \mu_2 \sim \mathcal{Be}(d_{N_1}, d_{N_2}),$$

kde

$$\begin{aligned}
d_{N_1} &= d_1 + \sum_{n=1}^N \delta(c_n, 1) \\
d_{N_2} &= d_2 + \sum_{n=1}^N \delta(c_n, 2)
\end{aligned}$$

$$\begin{aligned}
p(\mu_1) p(\mu_2) \prod_{n=1}^N p(x_n | \mu_1, \mu_2, c_n) &\propto \\
\exp\left(\frac{-(\mu_1 - \bar{\mu}_1)^2}{2\bar{\sigma}_1^2}\right) \exp\left(\frac{-(\mu_2 - \bar{\mu}_2)^2}{2\bar{\sigma}_2^2}\right) \exp\left(\sum_{n=1}^N \delta(1, c_n) \frac{-(x_n - \mu_1)^2}{2\sigma_1^2}\right) \\
\exp\left(\sum_{n=1}^N \delta(2, c_n) \frac{-(x_n - \mu_2)^2}{2\sigma_2^2}\right)
\end{aligned}$$

Obdobným způsobem jako u příkladu (2) dostáváme

$$\mu_1, \mu_2 | x, c, \alpha_1 \sim \mathcal{N}(\mu_1 | \mu_{N_1}, \sigma_{N_1}^2) \otimes \mathcal{N}(\mu_2 | \mu_{N_2}, \sigma_{N_2}^2)$$

s novými parametry $\mu_{N_1}, \mu_{N_2}, \sigma_{N_1}^2, \sigma_{N_2}^2$.

$$\begin{aligned}
\sigma_{N_1}^2 &= \frac{\sigma_1^2 \bar{\sigma}_1^2}{\sigma_1^2 + \bar{\sigma}_1^2 \sum_{n=1}^N \delta(1, c_n)} = \frac{1}{\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \delta(1, c_n)}{\sigma_1^2}} \\
\mu_{N_1} &= \frac{\sigma_1^2 \bar{\mu}_1 + \bar{\sigma}_1^2 \sum_{n=1}^N \delta(1, c_n) x_n}{\sigma_1^2 \bar{\sigma}_1^2} = \sigma_{N_1}^2 \left(\frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \delta(1, c_n) x_n}{\sigma_1^2} \right) \\
\sigma_{N_2}^2 &= \frac{1}{\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \delta(2, c_n)}{\sigma_2^2}} \\
\mu_{N_2} &= \sigma_{N_2}^2 \left(\frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \delta(2, c_n) x_n}{\sigma_2^2} \right)
\end{aligned}$$

Plně podmíněné rozdělení $p(\mu_1, \mu_2, \alpha_1 | x, c)$ je tedy dáno součinem beta rozdělení a dvou normálních rozdělení:

$$\mathcal{B}e(d_{N_1}, d_{N_2}) \otimes \mathcal{N}(\mu_1 | \mu_{N_1}, \sigma_{N_1}^2) \otimes \mathcal{N}(\mu_2 | \mu_{N_2}, \sigma_{N_2}^2)$$

Vyjádříme plně podmíněné rozdělení $p(c_n = 1 | x_n, \mu_1, \mu_2, \alpha_1)$ a $p(c_n = 2 | x_n, \mu_1, \mu_2, \alpha_1)$

$$\begin{aligned}
p(c_n = 1 | x_n, \mu_1, \mu_2, \alpha_1) &= \frac{\alpha_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(x_n - \mu_1)^2}{2\sigma_1^2}\right)}{\alpha_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(x_n - \mu_1)^2}{2\sigma_1^2}\right) + (1 - \alpha_1) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(x_n - \mu_2)^2}{2\sigma_2^2}\right)} \\
&\stackrel{\text{ozn.}}{=} \frac{\kappa_1}{\kappa_1 + \kappa_2} \\
p(c_n = 2 | x_n, \mu_1, \mu_2, \alpha_1) &= \frac{(1 - \alpha_1) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(x_n - \mu_2)^2}{2\sigma_2^2}\right)}{\alpha_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(x_n - \mu_1)^2}{2\sigma_1^2}\right) + (1 - \alpha_1) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(x_n - \mu_2)^2}{2\sigma_2^2}\right)} \\
&\stackrel{\text{ozn.}}{=} \frac{\kappa_2}{\kappa_1 + \kappa_2}
\end{aligned}$$

Formulujeme-li plně podmíněné rozdělení ve tvaru exponenciální rodiny, dostáváme:

$$\begin{aligned}
p(\mu_1, \mu_2, \alpha_1 | x, c) &= \\
&= \frac{1}{\sqrt{2\pi\alpha_1(1-\alpha_1)}} \exp\left(\left(\frac{\mu_{N_1}}{\sigma_{N_1}^2}, \frac{-1}{2\sigma_{N_1}^2}, \frac{\mu_{N_2}}{\sigma_{N_2}^2}, \frac{-1}{2\sigma_{N_2}^2}, d_{N_1}, d_{N_2} \right)^T \left(\mu_1, \mu_1^2, \mu_2, \mu_2^2, \ln \alpha_1, \ln(1 - \alpha_1) \right) \right. \\
&\quad \left. - \left(\frac{\mu_{N_1}^2}{2\sigma_{N_1}^2} + \ln(\sigma_{N_2}) + \frac{\mu_{N_2}^2}{2\sigma_{N_2}^2} + \ln(\sigma_{N_2}) + \ln \gamma(d_{N_1}) + \ln \gamma(d_{N_2}) - \ln \gamma(d_{N_1} + d_{N_2}) \right) \right)
\end{aligned}$$

$$\begin{aligned}
p(c_n|x_n, \alpha_1, \mu_1, \mu_2) &= \\
&= \exp \left(\left(\ln \frac{\kappa_1}{\kappa_1 + \kappa_2}, \ln \frac{\kappa_2}{\kappa_1 + \kappa_2} \right)^T (\delta(c_n, 1), \delta(c_n, 2)) \right) = \\
&= \exp \left(((\ln \kappa_1, \ln \kappa_2) (\delta(c_n, 1), \delta(c_n, 2)))^T - (\ln(\kappa_1 + \kappa_2), \ln \lambda(\kappa_1 + \kappa_2)) (\delta(c_n, 1), \delta(c_n, 2)) \right) = \\
&= \exp \left((\ln \kappa_1, \ln \kappa_2) (\delta(c_n, 1), \delta(c_n, 2))^T - \ln(\kappa_1 + \kappa_2) \right)
\end{aligned}$$

Variační rozdělení pro vektor globálních parametrů μ_1, μ_2, α_1 volíme (viz 4.2) ve stejném tvaru jako plně podmíněné rozdělení pomocí parametrizace $(\mu_{N_1}^*, (\sigma_{N_1}^*)^2, \mu_{N_2}^*, (\sigma_{N_2}^*)^2, d_{N_1}^*, d_{N_2}^*)$. Označíme-li přirozený parametr rozdělení λ , dostaneme vztah mezi λ a $(\mu_{N_1}^*, (\sigma_{N_1}^*)^2, \mu_{N_2}^*, (\sigma_{N_2}^*)^2, d_{N_1}^*, d_{N_2}^*)$.

$$q(\mu_1, \mu_2, \alpha_1 | \lambda) \propto \exp \left(\lambda^T (\mu_1, \mu_1^2, \mu_2, \mu_2^2, \ln \alpha_1, \ln(1 - \alpha_1)) \right)$$

$$\begin{aligned}
q(\mu_1, \mu_2, \alpha_1 | \mu_{N_1}^*, (\sigma_{N_1}^*)^2, \mu_{N_2}^*, (\sigma_{N_2}^*)^2, d_{N_1}^*, d_{N_2}^*) &= \\
&= \frac{1}{\sqrt{2\pi} \alpha_1 (1 - \alpha_1)} \exp \left(\left(\frac{\mu_{N_1}^*}{(\sigma_{N_1}^*)^2}, \frac{-1}{2(\sigma_{N_1}^*)^2}, \frac{\mu_{N_2}^*}{(\sigma_{N_2}^*)^2}, \frac{-1}{2(\sigma_{N_2}^*)^2}, d_{N_1}^*, d_{N_2}^* \right)^T \right. \\
&\quad \left. (\mu_1, \mu_1^2, \mu_2, \mu_2^2, \ln \alpha_1, \ln(1 - \alpha_1)) \right) \\
&- \left(\frac{(\mu_{N_1}^*)^2}{(2\sigma_{N_1}^*)^2} + \ln(\sigma_{N_2}^*) + \frac{(\mu_{N_2}^*)^2}{(2\sigma_{N_2}^*)^2} + \ln(\sigma_{N_2}^*) + \ln \Gamma(d_{N_1}^*) + \ln \Gamma(d_{N_2}^*) - \ln \Gamma(d_{N_1}^* + d_{N_2}^*) \right).
\end{aligned}$$

Pro lokální parametry c_n zapíšeme variační rozdělení pomocí parametrů $(\theta_n, 1 - \theta_n)$, které vzniknou transformací přirozeného parametru ϕ

$$q(c_n | \theta_n) = \begin{cases} \theta_n & \text{pro } c_n = 1 \\ 1 - \theta_n & \text{pro } c_n = 2 \end{cases}$$

Optimální parametry λ a ϕ_n počítáme iteračně ze středních hodnot přirozených parametrů plně podmíněného rozdělení (viz 4.2)

$$\begin{aligned}
\lambda &= E_q \left[\frac{\mu_{N_1}}{\sigma_{N_1}^2}, \frac{-1}{2\sigma_{N_1}^2}, \frac{\mu_{N_2}}{\sigma_{N_2}^2}, \frac{-1}{2\sigma_{N_2}^2}, d_{N_1}, d_{N_2} \right]^T \\
\phi_n &= E_q [\ln \kappa_1, \ln \kappa_2]^T
\end{aligned}$$

Hodnotu $\phi^{[i+1]}$ v $i + 1$ získáme pomocí přechodí iterace $\lambda^{[i]}$ a $\lambda^{[i+1]}$ dostaneme z $\phi^{[i+1]}$. Pro přehlednost vynecháváme v textu indexy $[i], [i + 1]$.

Spočítáme tedy příslušné střední hodnoty

$$\begin{aligned}
E_q \left[\frac{\mu_{N_1}}{\sigma_{N_1}^2} \right] &= E_q \left[\frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \delta(1, c_n) x_n}{\sigma_1^2} \right] \\
&= \frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N E_q [\delta(1, c_n) x_n]}{\sigma_1^2} \\
&= \frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \sum_{k=1}^2 \delta(1, k) q(c_n = k) x_n}{\sigma_1^2} \\
&= \frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \theta_n x_n}{\sigma_1^2}
\end{aligned}$$

$$\begin{aligned}
E_q \left[\frac{-1}{2\sigma_{N_1}^2} \right] &= E_q \left[\frac{-1}{2} \left(\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \delta(1, c_n)}{\sigma_1^2} \right) \right] \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N E_q [\delta(1, c_n)]}{\sigma_1^2} \right) \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \sum_{k=1}^2 \delta(1, k) q(c_n = k)}{\sigma_1^2} \right) \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \theta_n}{\sigma_1^2} \right)
\end{aligned}$$

$$\begin{aligned}
E_q \left[\frac{\mu_{N_2}}{\sigma_{N_2}^2} \right] &= E_q \left[\frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \delta(2, c_n) x_n}{\sigma_1^2} \right] \\
&= \frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N E_q [\delta(2, c_n) x_n]}{\sigma_2^2} \\
&= \frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \sum_{k=1}^2 \delta(2, k) q(c_n = k) x_n}{\sigma_2^2} \\
&= \frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N (1 - \theta_n) x_n}{\sigma_2^2} \\
E_q \left[\frac{-1}{2\sigma_{N_2}^2} \right] &= E_q \left[\frac{-1}{2} \left(\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \delta(2, c_n)}{\sigma_2^2} \right) \right] \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N E_q [\delta(2, c_n)]}{\sigma_2^2} \right) \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N \sum_{k=1}^2 \delta(2, k) q(c_n = k)}{\sigma_2^2} \right) \\
&= \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N (1 - \theta_n)}{\sigma_2^2} \right) \\
E_q [d_{N_1}] &= E_q \left[d_1 + \sum_{n=1}^N \delta(1, c_n) \right] \\
&= d_1 + \sum_{n=1}^N \theta_n \\
E_q [d_{N_2}] &= E_q \left[d_2 + \sum_{n=1}^N \delta(2, c_n) \right] \\
&= d_2 + \sum_{n=1}^N (1 - \theta_n)
\end{aligned}$$

$$\begin{aligned}
E_q[\ln \kappa_1] &= E_q \left[\left(\ln \alpha_1 - \ln \sqrt{2\pi\sigma_1^2} \right) + \frac{-(x_n - \mu_1)^2}{2\sigma_1^2} \right] \\
&= \left(E_q[\ln \alpha_1] - \ln \sqrt{2\pi\sigma_1^2} \right) + \frac{E_q[-(x_n - \mu_1)^2]}{2\sigma_1^2}
\end{aligned}$$

Střední hodnotou $\ln \alpha_1$ lze vyjádřit pomocí digamma funkce² $E_q[\ln \alpha_1] = \psi(d_{N_1}^*) - \psi(d_{N_1}^* + d_{N_2}^*)$.

$$\begin{aligned}
E_q[-(x_n - \mu_1)^2] &= E_q[-x_n^2 + 2x_n\mu_1 - \mu_1^2] \\
&= -x_n^2 + 2x_n\mu_{N_1}^* - ((\sigma_{N_1}^*)^2 + (\mu_{N_1}^*)^2)
\end{aligned}$$

Dostáváme tedy

$$E_q[\ln \kappa_1] = \left(\psi(d_{N_1}^*) - \psi(d_{N_1}^* + d_{N_2}^*) - \ln \sqrt{2\pi\sigma_1^2} \right) + \frac{-x_n^2 + 2x_n\mu_{N_1}^* - (\sigma_{N_1}^*)^2 - (\mu_{N_1}^*)^2}{2\sigma_1^2}.$$

Obdobným způsobem dostaneme střední hodnotu $\ln \kappa_2$

$$E_q[\ln \kappa_2] = \left(\psi(d_{N_2}^*) - \psi(d_{N_1}^* + d_{N_2}^*) - \ln \sqrt{2\pi\sigma_2^2} \right) + \frac{-x_n^2 + 2x_n\mu_{N_2}^* - (\sigma_{N_2}^*)^2 - (\mu_{N_2}^*)^2}{2\sigma_2^2}$$

Tímto jsme získali přibližné hodnoty parametrů λ a ϕ_n variační aproximace q

$$\begin{aligned}
\lambda &= \left(\frac{\bar{\mu}_1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \theta_n x_n}{\sigma_1^2}, \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_1^2} + \frac{\sum_{n=1}^N \theta_n}{\sigma_1^2} \right), \frac{\bar{\mu}_2}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N (1 - \theta_n) x_n}{\sigma_2^2}, \frac{-1}{2} \left(\frac{1}{\bar{\sigma}_2^2} + \frac{\sum_{n=1}^N (1 - \theta_n)}{\sigma_2^2} \right), \right. \\
&\quad \left. d_1 + \sum_{n=1}^N \theta_n, d_2 + \sum_{n=1}^N (1 - \theta_n) \right)
\end{aligned}$$

$$\begin{aligned}
\phi_n &= \left(\left(\psi(d_{N_1}^*) - \psi(d_{N_1}^* + d_{N_2}^*) - \ln \sqrt{2\pi\sigma_1^2} \right) + \frac{-x_n^2 + 2x_n\mu_{N_1}^* - (\sigma_{N_1}^*)^2 - (\mu_{N_1}^*)^2}{2\sigma_1^2}, \right. \\
&\quad \left. \left(\psi(d_{N_2}^*) - \psi(d_{N_1}^* + d_{N_2}^*) - \ln \sqrt{2\pi\sigma_2^2} \right) + \frac{-x_n^2 + 2x_n\mu_{N_2}^* - (\sigma_{N_2}^*)^2 - (\mu_{N_2}^*)^2}{2\sigma_2^2} \right)
\end{aligned}$$

²Digamma funkce je definována jako derivace logaritmu gamma funkce: $\psi(x) = \frac{d}{dx} \ln(\Gamma(x))$

Nyní převedeme přirozené parametry λ, ϕ na parametry běžné. Rozepíšeme si plně podmíněné variační rozdělení s běžnými parametry

$$q\left(\mu_1, \mu_2, \alpha_1 | \mu_{N_1}^*, (\sigma_{N_1}^*)^2, \mu_{N_2}^*, (\sigma_{N_2}^*)^2, d_{N_1}^*, d_{N_2}^*\right) \text{ na součin plně podmíněných rozdělení } \\ q\left(\mu_1 | \mu_{N_1}^*, (\sigma_{N_1}^*)^2\right) q\left(\mu_2 | \mu_{N_2}^*, (\sigma_{N_2}^*)^2\right) q\left(\alpha_1 | d_{N_1}^*, d_{N_2}^*\right).$$

Plně podmíněné rozdělení s parametrem λ si taktéž rozložíme na součin:

$$q(\mu_1, \mu_2, \alpha_1 | \lambda) = q(\mu_1 | \lambda_1, \lambda_2) q(\mu_2 | \lambda_3, \lambda_4) q(\alpha_1 | \lambda_5, \lambda_6). \text{ Tato rozdělení pak zapíšeme ve tvaru } \\ \text{exponenciálních rodn. Z parametrů } \lambda \text{ dostaneme transformací hledané parametry } \mu_{N_1}^*, \mu_{N_2}^*, \\ (\sigma_{N_1}^*)^2, (\sigma_{N_2}^*)^2, d_{N_1}^*, d_{N_2}^*.$$

$$q\left(\mu_1 | \mu_{N_1}^*, (\sigma_{N_1}^*)^2\right) = \frac{1}{\sqrt{2\pi} (\sigma_{N_1}^*)^2} \exp\left(\frac{-\left(\mu_1 - \mu_{N_1}^*\right)^2}{2 (\sigma_{N_1}^*)^2}\right) \\ = \frac{1}{\sqrt{2}} \exp\left(\left(\left(\frac{\mu_{N_1}^*}{(\sigma_{N_1}^*)^2}, \frac{-1}{2 (\sigma_{N_1}^*)^2}\right)^T (\mu_1, \mu_1^2) - \left(\frac{(\mu_{N_1}^*)^2}{2 (\sigma_{N_1}^*)^2} + \ln \sigma_{N_1}^*\right)\right)\right)$$

$$\lambda_1 = \frac{\mu_{N_1}^*}{(\sigma_{N_1}^*)^2} \quad \lambda_2 = \frac{-1}{2 (\sigma_{N_1}^*)^2} \quad \rightarrow \quad \begin{aligned} \mu_{N_1}^* &= \frac{-\lambda_1}{2\lambda_2} \\ (\sigma_{N_1}^*)^2 &= \frac{-1}{2\lambda_2} \end{aligned}$$

$$\lambda_3 = \frac{\mu_{N_2}^*}{(\sigma_{N_2}^*)^2} \quad \lambda_4 = \frac{-1}{2 (\sigma_{N_2}^*)^2} \quad \rightarrow \quad \begin{aligned} \mu_{N_2}^* &= \frac{-\lambda_3}{2\lambda_4} \\ (\sigma_{N_2}^*)^2 &= \frac{-1}{2\lambda_4} \end{aligned}$$

$$q(\alpha_1 | d_{N_1}^*, d_{N_2}^*) = \frac{\alpha_1^{d_{N_1}^*} (1 - \alpha_1)^{d_{N_2}^* - 1}}{B(d_{N_1}^*, d_{N_2}^*)} \\ = \exp\left(\ln \frac{\alpha_1^{d_{N_1}^* - 1} (1 - \alpha_1)^{d_{N_2}^* - 1}}{B(d_{N_1}^*, d_{N_2}^*)}\right) \\ = \frac{1}{\alpha_1 (1 - \alpha_1)} \exp(d_{N_1}^* \ln \alpha_1 + d_{N_2}^* \ln (1 - \alpha_1) - \ln B(d_{N_1}^*, d_{N_2}^*)) \\ = \frac{1}{\alpha_1 (1 - \alpha_1)} \exp\left((d_{N_1}^*, d_{N_2}^*)^T (\ln \alpha_1, \ln (1 - \alpha_1)) - \ln B(d_{N_1}^*, d_{N_2}^*)\right)$$

$$\lambda_5 = d_{N_1}^* \rightarrow d_{N_1}^* = \lambda_5$$

$$\lambda_6 = d_{N_2}^* \rightarrow d_{N_2}^* = \lambda_6$$

Rozdělení $q(c_n|\theta_n, 1 - \theta_n)$ si taktéž zapíšeme v exponenciálním tvaru a následně z něj pomocí parametru ϕ_n získáme parametry θ_n a $1 - \theta_n$.

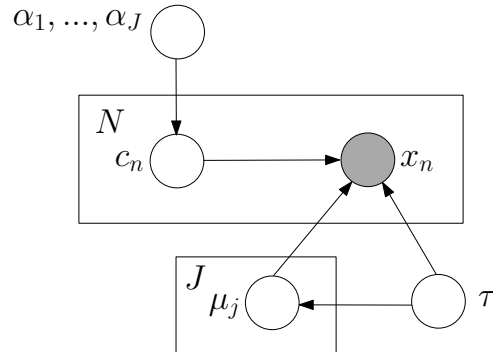
$$q(c_n|\theta_n, 1 - \theta_n) = \begin{cases} \theta_n & \text{pro } c_n = 1 \\ 1 - \theta_n & \text{pro } c_n = 2 \end{cases}$$

$$\begin{aligned} q(c_n|\theta_n, 1 - \theta_n) &= \exp(\ln \theta_n \delta(c_n, 1) + \ln(1 - \theta_n) \delta(c_n, 2)) \\ &= \exp\left((\ln \theta_n, \ln(1 - \theta_n))^T (\delta(c_n, 1), \delta(c_n, 2))\right) \end{aligned}$$

$$\begin{aligned} \phi_{n1} = \ln \theta_n &\rightarrow \theta_n = \exp \phi_1 \\ \phi_{n2} = \ln(1 - \theta_n) &\rightarrow 1 - \theta_n = \exp \phi_2 \end{aligned}$$

Numerický experiment pro tento model je v kapitole 5, jako příklad 3.

4.4 Odhad parametrů směšového modelu s n komponentami



Obrázek 5: Grafový směšový model s J komponentami

Budeme uvažovat směšový model s J gaussovskými komponentami se společnou přesností τ . Skrytými veličinami jsou zde střední hodnoty $\mu = (\mu_1, \dots, \mu_J)$, přesnost τ , váhy komponent $\alpha = (\alpha_1, \dots, \alpha_J)$ a čísla komponent c_n určující, z které komponenty pozorované x_n pochází. Grafový model pro tuto směs lze vidět na obrázku 5.

Data x_n budou pocházet z normálního rozdělení:

$$\begin{aligned} x_n | \mu, \alpha, \tau, c_n &\sim \mathcal{N}(x_n | \mu_{c_n}, \sigma_{c_n}^2) = \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp\left(-\frac{\tau}{2} (x_n - \mu_{c_n})^2\right) = \\ &= \prod_{j=1}^J \left(\frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp\left(-\frac{\tau}{2} (x_n - \mu_{c_n})^2\right) \right)^{\delta(c_n, j)} \end{aligned}$$

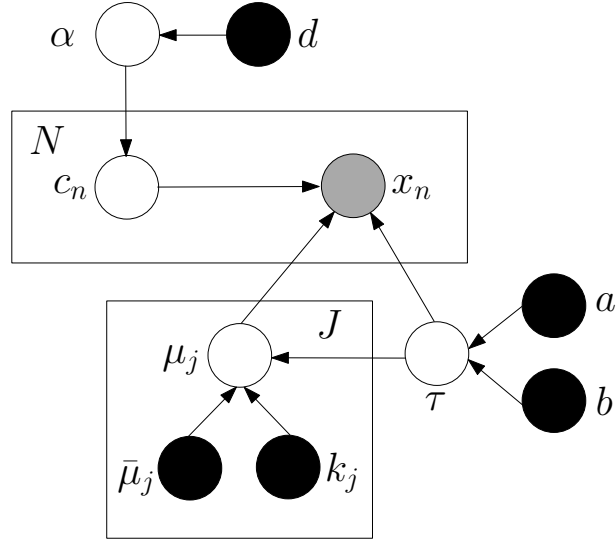
Sdruženou pravděpodobnost vyjádříme s pomocí podmíněných nezávislostí:

$$\begin{aligned} p(x, c, \mu, \alpha, \tau) &= p(\mu, \alpha, \tau) \prod_{n=1}^N p(x_n, c_n | \mu, \alpha, \tau) = \\ &= p(\mu, \alpha, \tau) \prod_{n=1}^N p(c_n | \alpha) p(x_n | \mu, \tau, c_n) = \\ &= p(\mu, \alpha, \tau) \prod_{n=1}^N \alpha_{c_n} \mathcal{N}(x_n | \mu_{c_n}, \tau^{-1}) \end{aligned}$$

Váhy α pocházejí z dirichletova rozdělení, rozdělení dvojice μ, τ je zobecněním normálního-gamma rozdělení, kde μ_1, \dots, μ_J jsou vzájemně podmíněně nezávislé za podmínky τ . Hodnoty c_n jsou z diskrétního rozdělení:

$$\begin{aligned} (\mu, \tau) &\sim \mathcal{NG} \\ \tau &\sim \mathcal{G}(a, b), p(\tau | a, b) = b^a \tau^{a-1} \exp(-b\tau) \frac{1}{\Gamma(a)} \\ \mu_j | \tau &\sim \mathcal{N}(\bar{\mu}_j, (k_j \tau)^{-1}), p(\mu_j | \bar{\mu}_j, (k_j \tau)^{-1}) = \frac{1}{\sqrt{2\pi}} k_j^{1/2} \tau^{1/2} \exp\left(-\frac{\tau k_j}{2} (\mu_j - \bar{\mu}_j)^2\right) \\ \alpha &\sim \mathcal{Dir}(d_1, \dots, d_J), p(\alpha | d_1, \dots, d_J) = \frac{1}{B(d_1, \dots, d_J)} \prod_{j=1}^J \alpha_j^{d_j-1} \\ c_n | \alpha &\sim D(\alpha_1, \dots, \alpha_J), p(c_n | \alpha_1, \dots, \alpha_J) = \alpha_{c_n} = \prod_{j=1}^J \alpha_j^{\delta(c_n, j)} \end{aligned}$$

Grafová reprezentace Bayesovského modelu je na obrázku 6.



Obrázek 6: Grafová reprezentace Bayesovského směšového modelu s J komponentami

Nejdříve si vyjádříme plně podmíněné rozdělení $p(\mu, \tau, \alpha | x, c)$, které je dáno jako součin $p(\alpha) \prod_{n=1}^N p(c_n | \alpha)$ a $p(\mu, \tau | x, c)$.

$$\begin{aligned}
 p(\alpha) \prod_{n=1}^N p(c_n | \alpha) &= \frac{1}{B(d_1, \dots, d_J)} \prod_{j=1}^J \alpha_j^{d_j-1} \prod_{j=1}^J \alpha_j^{\sum_{n=1}^N \delta(c_n, j)} = \\
 &= \frac{1}{B(d_1, \dots, d_J)} \prod_{j=1}^J \alpha_j^{d_j + \sum_{n=1}^N \delta(c_n, j) - 1} \\
 &\propto \prod_{j=1}^J \alpha_j^{d_j + \sum_{n=1}^N \delta(c_n, j) - 1}
 \end{aligned}$$

$$\alpha | x, c, \mu, \tau \sim \text{Dir}(d_{N_1}, \dots, d_{N_J})$$

$$d_{N_j} = d_j + \sum_{n=1}^N \delta(c_n, j)$$

$$\begin{aligned}
p(\mu, \tau | x, c) &\propto p(\mu, \tau) p(x | c, \mu, \tau) \propto \prod_{j=1}^J \mathcal{N}(\bar{\mu}_j, (k_j \tau)^{-1}) \prod_{n=1}^N \mathcal{N}(x_n | \mu_{c_n}, \tau^{-1}) \mathcal{G}(a, b) \propto \\
&\propto \prod_{j=1}^J k_j^{1/2} \left(\frac{1}{\sqrt{2\pi}} \tau^{1/2} \right)^J \exp \left(-\frac{\tau}{2} \sum_{j=1}^J k_j (\mu_j - \bar{\mu}_j)^2 \right) \left(\frac{1}{\sqrt{2\pi}} \tau^{1/2} \right)^N \\
&\exp \left(\sum_{j=1}^J \left(-\frac{\tau}{2} \sum_{n=1}^N (x_n - \mu_j)^2 \delta(c_n, j) \right) \right) \frac{b^a \tau^{a-1} \exp(-b\tau)}{\Gamma(a)} \propto \\
&\propto \tau^{J/2} \exp \left(-\frac{\tau}{2} \sum_{j=1}^J k_j (\mu_j - \bar{\mu}_j)^2 \right) \tau^{N/2} \exp \left(\sum_{j=1}^J \left(-\frac{\tau}{2} \sum_{n=1}^N (x_n - \mu_j)^2 \delta(c_n, j) \right) \right) \tau^{a-1} \exp(-b\tau) \propto \\
&\propto \tau^{J/2} \tau^{a+N/2-1} \exp(-b\tau) \exp \left(\sum_{j=1}^J \left(-\frac{\tau}{2} \left(k_j (\mu_j - \bar{\mu}_j)^2 + \sum_{n=1}^N (x_n - \mu_j)^2 \delta(c_n, j) \right) \right) \right)
\end{aligned} \tag{18}$$

Pro přehlednost zavedeme následující označení:

$$\begin{aligned}
\mathbb{C}_j &= \sum_{n=1}^N \delta(c_n, j) \\
\mathbb{X}_j &= \sum_{n=1}^N x_n \delta(c_n, j) \\
\mathbb{S}_j &= \sum_{n=1}^N x_n^2 \delta(c_n, j).
\end{aligned}$$

Pak sumu v exponenciále v 18 lze zapsat:

$$\begin{aligned}
&\sum_{j=1}^J \left(k_j (\mu_j - \bar{\mu}_j)^2 + \sum_{n=1}^N (x_n - \mu_j)^2 \delta(c_n, j) \right) = \\
&= \sum_{j=1}^J \left((k_j + \mathbb{C}_j) (\mu_j - \mu_{Nj})^2 + \frac{k_j \mathbb{C}_j \left(\frac{\mathbb{X}_j}{\mathbb{C}_j} - \bar{\mu}_j \right)^2}{k_j + \mathbb{C}_j} + \left(\sum_{n=1}^N \delta(c_n, j) \left(x_n - \frac{\mathbb{X}_j}{\mathbb{C}_j} \right)^2 \right) \right),
\end{aligned}$$

kde

$$\mu_{Nj} = \frac{k_j \bar{\mu}_j + \mathbb{X}_j}{k_j + \mathbb{C}_j}$$

$$\begin{aligned}
p(\mu, \tau | x, c) &\propto \\
&\tau^{\frac{J}{2}} \tau^{a + \frac{N}{2} - 1} \exp(-b\tau) \\
&\exp \sum_{j=1}^J \left(-\frac{\tau}{2} \right) \left((k_j + \mathbb{C}_j) (\mu_j - \mu_{Nj})^2 + \frac{k_j \mathbb{C}_j \left(\frac{\mathbb{X}_j}{\mathbb{C}_j} - \bar{\mu}_j \right)^2}{k_j + \mathbb{C}_j} + \left(\sum_{n=1}^N \delta(c_n, j) \left(x_n - \frac{\mathbb{X}_j}{\mathbb{C}_j} \right)^2 \right) \right) \propto \\
&\propto \left(\prod_{j=1}^J \tau^{1/2} \exp \left(-\frac{\tau (k_j + \mathbb{C}_j)}{2} \right) (\mu_j - \mu_{Nj})^2 \right) \tau^{a + N/2 - 1} \\
&\exp -\tau \left(b + \frac{1}{2} \sum_{j=1}^J \left(\frac{k_j \mathbb{C}_j \left(\frac{\mathbb{X}_j}{\mathbb{C}_j} - \bar{\mu}_j \right)^2}{k_j + \mathbb{C}_j} + \left(\sum_{n=1}^N \delta(c_n, j) \left(x_n - \frac{\mathbb{X}_j}{\mathbb{C}_j} \right)^2 \right) \right) \right) \\
&\propto \mathcal{G}(\tau | a_N, b_N) \prod_{j=1}^J \mathcal{N}(\mu_j | \mu_{Nj}, (k_{Nj} \tau)^{-1}),
\end{aligned}$$

kde

$$\begin{aligned}
k_{Nj} &= k_j + \mathbb{C}_j \\
a_N &= a + \frac{N}{2} \\
b_N &= b + \frac{1}{2} \sum_{j=1}^J \left(\frac{k_j \mathbb{C}_j \left(\frac{\mathbb{X}_j}{\mathbb{C}_j} - \bar{\mu}_j \right)^2}{k_j + \mathbb{C}_j} + \sum_{n=1}^N \delta(c_n, j) \left(x_n - \frac{\mathbb{X}_j}{\mathbb{C}_j} \right)^2 \right)
\end{aligned}$$

Plně podmíněné rozdělení $p(\mu, \tau, \alpha | x, c)$ se tedy rovná součinu dirichletova rozdělení, gamma rozdělení a normálních rozdělení:

$$\mathcal{Dir}(d_{N_1}, \dots, d_{N_J}) \otimes \mathcal{N}(\mu_1 | \mu_{N_1}, \tau) \otimes \dots \otimes \mathcal{N}(\mu_J | \mu_{N_J}, \tau) \otimes \mathcal{G}(\tau | a_N, b_N)$$

Nyní si vyjádříme plně podmíněné rozdělení $p(c_n = j | x_n, \mu, \tau, \alpha)$ pro j -tou komponentu:

$$\begin{aligned}
p(c_n = j | x_n, \mu, \tau, \alpha) &= \frac{\alpha_j \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp \left(-\frac{\tau}{2} (x_n - \mu_j)^2 \right)}{\sum_{\tilde{j}=1}^J \alpha_{\tilde{j}} \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp \left(-\frac{\tau}{2} (x_n - \mu_{\tilde{j}})^2 \right)} \\
&\stackrel{\text{ozn.}}{=} \frac{\kappa_j}{\sum_{\tilde{j}=1}^J \kappa_{\tilde{j}}}
\end{aligned}$$

Plně podmíněná rozdělení si opět vyjádříme ve tvaru exponenciálních rodin:

$$\begin{aligned}
p(\mu_1, \mu_2, \alpha_1 | x, c) &= \\
&= \frac{1}{(\sqrt{2\pi})^J} \frac{1}{\prod_{j=1}^J \alpha_j} \exp \left(\left(a_N - 1 + \frac{J}{2}, -b_N - \frac{1}{2} \sum_{j=1}^J k_{Nj} (\mu_{Nj})^2, k_{N1} \mu_{N1}, \dots, k_{NJ} \mu_{NJ}, \right. \right. \\
&\quad \left. \left. - \frac{k_{N1}}{2}, \dots, -\frac{k_{NJ}}{2}, d_{N1}, \dots, d_{NJ} \right)^T \left(\ln \tau, \tau, \tau \mu_1, \dots, \tau \mu_J, \tau \mu_1^2, \dots, \tau \mu_J^2, \ln \alpha_1, \dots, \ln \alpha_J \right) \right. \\
&\quad \left. - \left(\ln \Gamma(a_N) - a_N \ln(b_N) - \frac{1}{2} \sum_{j=1}^J \ln(k_{Nj}) + \sum_{j=1}^J \ln \Gamma(d_{Nj}) - \ln \Gamma \left(\sum_{j=1}^J d_{Nj} \right) \right) \right) \\
p(c_n | x_n, \alpha, \tau, \mu) &= \exp \left((\ln \kappa_1, \dots, \ln \kappa_J)^T (\delta(c_n, 1), \dots, \delta(c_n, J)) - \ln \left(\sum_{j=1}^J \kappa_j \right) \right)
\end{aligned}$$

V obdobném tvaru si zapíšeme variační rozdělení pro vektor globálních parametrů μ, τ, α pomocí parametrizace $(\mu_N^*, k_N^*, a_N^*, b_N^*, d_N^*)$ na kterou je transformován přirozený globální parametr λ :

$$\begin{aligned}
q(\mu_1, \mu_2, \alpha_1 | \mu_N^*, k_N^*, a_N^*, b_N^*, d_N^*) &= \\
&= \frac{1}{(\sqrt{2\pi})^J} \frac{1}{\prod_{j=1}^J \alpha_j} \exp \left(\left(a_N^* - 1 + \frac{J}{2}, -b_N^* - \frac{1}{2} \sum_{j=1}^J k_{Nj}^* (\mu_{Nj}^*)^2, k_{N1}^* \mu_{N1}^*, \dots, k_{NJ}^* \mu_{NJ}^*, \right. \right. \\
&\quad \left. \left. - \frac{k_{N1}^*}{2}, \dots, -\frac{k_{NJ}^*}{2}, d_{N1}^*, \dots, d_{NJ}^* \right)^T \left(\ln \tau, \tau, \tau \mu_1, \dots, \tau \mu_J, \tau \mu_1^2, \dots, \tau \mu_J^2, \ln \alpha_1, \dots, \ln \alpha_J \right) \right. \\
&\quad \left. - \left(\ln \Gamma(a_N^*) - a_N^* \ln(b_N^*) - \frac{1}{2} \sum_{j=1}^J \ln(k_{Nj}^*) + \sum_{j=1}^J \ln \Gamma(d_{Nj}^*) - \ln \Gamma \left(\sum_{j=1}^J d_{Nj}^* \right) \right) \right)
\end{aligned}$$

Variační rozdělení pro lokální parametry c_n zapíšeme pomocí parametrizace $(\theta_{n1}, \dots, \theta_{nJ})$ získané transformací přirozeného parametru ϕ

$$q(c_n = j) = \theta_{nj},$$

kde platí

$$\sum_{j=1}^J \theta_{nj} = 1.$$

Ze středních hodnot přirozených parametrů plně podmíněných rozdělání se nyní iteračně spočítají nové aproximace parametrů λ a ϕ_n .

$$\lambda = E_q \left[a_N - 1 + \frac{J}{2}, -b_N - \frac{1}{2} \sum_{j=1}^J k_{Nj} (\mu_{Nj})^2, k_{N1}\mu_{N1}, \dots, k_{NJ}\mu_{NJ}, -\frac{k_{N1}}{2}, \dots, -\frac{k_{NJ}}{2}, d_{N1}, \dots, d_{NJ} \right]^T$$

$$\phi_n = E_q [\ln \kappa_1, \dots, \ln \kappa_J]^T$$

Vypočteme požadované střední hodnoty:

$$E_q \left[a_N - 1 + \frac{J}{2} \right] = E_q \left[a - 1 + \frac{J}{2} + \frac{N}{2} \right] = a - 1 + \frac{J}{2} + \frac{N}{2}$$

$$\begin{aligned} E_q \left[-b_N - \frac{1}{2} \sum_{j=1}^J k_{Nj} (\mu_{Nj})^2 \right] &= \\ &= -b - \frac{1}{2} E_q \left[\sum_{j=1}^J \left(\left(\sum_{n=1}^N \delta(c_n, j) \left(x_n - \frac{\mathbb{X}_j}{\mathbb{C}_j} \right)^2 \right) + \frac{k_j \mathbb{C}_j \left(\frac{\mathbb{X}_j}{\mathbb{C}_j} - \bar{\mu}_j \right)^2}{k_j + \mathbb{C}_j} + \frac{(k_j \bar{\mu}_j + \mathbb{X}_j)^2}{k_j + \mathbb{C}_j} \right) \right] = \dots = \\ &= -b - \frac{1}{2} E_q \left[\sum_{j=1}^J (\mathbb{S}_j + k_j \bar{\mu}_j^2) \right] = -b - \frac{1}{2} \left(\sum_{n=1}^N x_n^2 + \sum_{j=1}^J k_j \bar{\mu}_j^2 \right) \end{aligned}$$

$$E_q [k_{Nj}\mu_{Nj}] = E_q \left[k_j \bar{\mu}_j + \sum_{n=1}^N x_n \delta(c_n, j) \right] = k_j \bar{\mu}_j + \sum_{n=1}^N x_n \theta_{nj}$$

$$E_q \left[-\frac{k_{Nj}}{2} \right] = E_q \left[-\frac{k_j + \sum_{n=1}^N \delta(c_n, j)}{2} \right] = -\frac{1}{2} \left(k_j + \sum_{n=1}^N \theta_{nj} \right)$$

$$\begin{aligned}
E_q [\ln \kappa_j] &= E_q \left[\ln \left(\alpha_j \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp \left(-\frac{\tau}{2} (x_n - \mu_j)^2 \right) \right) \right] = \\
&= E_q [\ln \alpha_j] - E_q [\ln \sqrt{2\pi}] + E_q \left[\frac{1}{2} \ln \tau \right] + E_q \left[-\frac{\tau}{2} (x_n - \mu_j)^2 \right] = \\
&= \psi(d_{Nj}^*) - \psi \left(\sum_{\tilde{j}=1}^J d_{Nj}^* \right) - \ln \sqrt{2\pi} + \frac{1}{2} (\psi(a_N^*) - \ln(b_N^*)) + \\
&\quad \left(-\frac{1}{2} \frac{a_N^*}{b_N^*} x_n^2 + \frac{a_N^*}{b_N^*} x_n \mu_{Nj}^* - \frac{1}{2} \left(\frac{1}{k_{Nj}^*} + \frac{a_N^*}{b_N^*} (\mu_{Nj}^*)^2 \right) \right)
\end{aligned}$$

Z vypočítaných středních hodnot dostáváme globální variační parametry λ a lokální variační parametry ϕ_{nj} :

$$\begin{aligned}
\lambda &= \left(a - 1 + \frac{J}{2} + \frac{N}{2}, -b - \frac{1}{2} \left(\sum_{n=1}^N x_n^2 + \sum_{j=1}^J k_j \bar{\mu}_j^2 \right), k_1 \bar{\mu}_1 + \sum_{n=1}^N x_n \theta_{n1}, \dots, k_J \bar{\mu}_J + \sum_{n=1}^N x_n \theta_{nJ}, \right. \\
&\quad \left. - \frac{1}{2} \left(k_1 + \sum_{n=1}^N \theta_{n1} \right), \dots, -\frac{1}{2} \left(k_J + \sum_{n=1}^N \theta_{nJ} \right) \right)
\end{aligned}$$

$$\begin{aligned}
\phi_n &= \left(\psi(d_{N1}^*) - \psi \left(\sum_{\tilde{j}=1}^J d_{Nj}^* \right) - \ln \sqrt{2\pi} + \frac{1}{2} (\psi(a_N^*) - \ln(b_N^*)) + \right. \\
&\quad \left(-\frac{1}{2} \frac{a_N^*}{b_N^*} x_n^2 + \frac{a_N^*}{b_N^*} x_n \mu_{N1}^* - \frac{1}{2} \left(\frac{1}{k_{N1}^*} + \frac{a_N^*}{b_N^*} \mu_{N1}^{*2} \right) \right), \dots, \psi(d_{NJ}^*) - \psi \left(\sum_{\tilde{j}=1}^J d_{Nj}^* \right) - \ln \sqrt{2\pi} + \\
&\quad \left. \frac{1}{2} (\psi(a_N^*) - \ln(b_N^*)) + \left(-\frac{1}{2} \frac{a_N^*}{b_N^*} x_n^2 + \frac{a_N^*}{b_N^*} x_n \mu_{NJ}^* - \frac{1}{2} \left(\frac{1}{k_{NJ}^*} + \frac{a_N^*}{b_N^*} (\mu_{NJ}^*)^2 \right) \right) \right)
\end{aligned}$$

Plně podmíněné variační rozdělení pro veličiny μ, α, τ lze parametrizovat buďto pomocí přirozeného globálního parametru λ nebo pomocí vektoru parametrů $(\mu_N^*, k_N^*, a_N^*, b_N^*, d_N^*)$. Pomocí transformace získáme z parametru λ parametry $\mu_N^*, k_N^*, a_N^*, b_N^*, d_N^*$.

$$\lambda = \left(a_N^* - 1 + \frac{J}{2}, -b_N^* - \frac{1}{2} \sum_{j=1}^J k_{Nj}^* (\mu_{Nj}^*)^2, k_{N1}^* \mu_{N1}^*, \dots, k_{NJ}^* \mu_{NJ}^*, -\frac{k_{N1}^*}{2}, \dots, -\frac{k_{NJ}^*}{2}, d_{N1}^*, \dots, d_{NJ}^* \right)^T$$

Pro $j = 1, \dots, J$:

$$\begin{aligned}
\lambda_1 &= a_N^* - 1 + \frac{J}{2} & a_N^* &= \lambda_1 + 1 - \frac{J}{2} \\
\lambda_{2+J+j} &= -\frac{k_{Nj}^*}{2} & k_{Nj}^* &= -2\lambda_{2+J+j} \\
\lambda_{2+j} &= k_{Nj}^* \mu_{Nj}^* = -2\lambda_{2+J+j} \mu_{Nj}^* & \longrightarrow & \mu_{Nj}^* = -\frac{\lambda_{2+j}}{2\lambda_{2+J+j}} \\
\lambda_2 &= -b_N^* - \frac{1}{2} \sum_{j=1}^J k_{Nj}^* (\mu_{Nj}^*)^2 = -b_N^* + \frac{1}{4} \sum_{j=1}^J \frac{(\lambda_{2+j})^2}{2\lambda_{2+J+j}} & b_N^* &= \frac{1}{4} \sum_{j=1}^J \frac{(\lambda_{2+j})^2}{\lambda_{2+J+j}} \\
\lambda_{2+2J+j} &= d_{Nj}^* & d_{Nj}^* &= \lambda_{2+2J+j}
\end{aligned}$$

Z plně podmíněného rozdělení pro lokální veličinu c_n si vyjádříme pomocí parametru ϕ_n parametry $(\theta_{nJ}, \dots, \theta_{n1})$.

Pro $j = 1, \dots, J$:

$$\begin{aligned}
\Phi_n &= (\ln \theta_{n1}, \dots, \ln \theta_{nJ}) \\
\Phi_{nj} &= \ln \theta_{nj} \rightarrow \theta_{nj} = \exp(\phi_{nj})
\end{aligned}$$

Numerické experimenty se nacházejí v kapitole 5. Jedná se o příklady 4 a 5.

5 Numerické experimenty

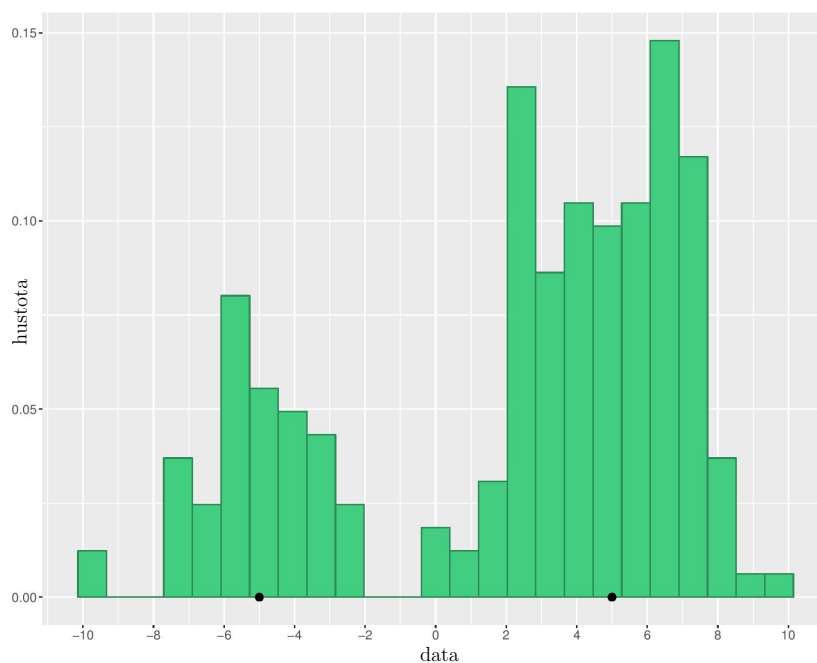
Příklad 3

Nyní budeme hledat, s pomocí algoritmu 1, variační parametry 2 komponentního směšového modelu se známým rozptylem, popsaného v kapitole 4.3. Budou použita data generovaná z 2 komponenti směsi pro počet dat $N = 200$.

Data byla generována za pomoci následujících parametrů:

$$\begin{aligned}\mu_1 &= -5 \\ \mu_2 &= 5 \\ \sigma_1 &= 2 \\ \sigma_2 &= 2 \\ \alpha_1 &= 0.3\end{aligned}\tag{19}$$

Z těchto parametrů byla nasimulována data, jejichž histogram je na obrázku 7.



Obrázek 7: Histogram nasimulovaných dat

Apriorní parametry byly zvoleny následovně:

$$\begin{aligned}\bar{\mu}_1 &= -3 \\ \bar{\mu}_2 &= 3 \\ \bar{\sigma}_1 &= 10 \\ \bar{\sigma}_2 &= 10 \\ d_1 &= 1 \\ d_2 &= 1\end{aligned}$$

Počáteční aproximace byla zvolena jako:

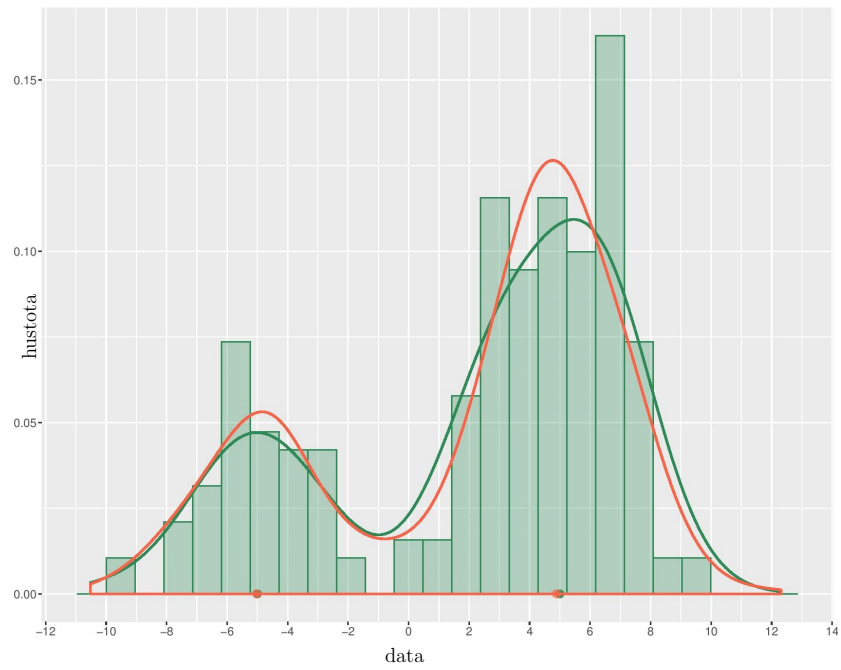
$$\begin{aligned}\mu_{N_1}^* &= -1 \\ \mu_{N_2}^* &= 1 \\ \sigma_{N_1}^* &= 1 \\ \sigma_{N_2}^* &= 1 \\ d_{N_1}^* &= 0.5 \\ d_{N_2}^* &= 0.5\end{aligned}\tag{20}$$

Po 9 iteracích jsme našli variační parametry. Ty jsou uvedeny v tabulce 1. Kritérium pro zastavení bylo zvoleno jako změna velikosti vektoru parametrů s prahem $\varepsilon = 10^{-4}$. Na obrázku 16 lze vidět histogram simulovaných dat (zeleně) generovaných z parametrů z (19) a skutečnou hustotu pravděpodobnosti (zelená čára). Oranžově je znázorněna aproximace prediktivní hustoty

$$p(x|x_1, \dots, x_{200}) = \int p(x|\alpha_1, \mu_1, \mu_2) p(\alpha_1, \mu_1, \mu_2|x_1, \dots, x_{200}),$$

kde místo skutečné posteriorní hustoty $p(\alpha_1, \mu_1, \mu_2|x_1, \dots, x_{200})$ je použita nalezená variační aproximace. Střední hodnoty simulovaných dat (zeleně) a predikovaných dat (oranžově) jsou znázorněny puntíky. Na obrázcích 9 až 14 vidíme, jak se mění parametry variační hustoty v závislosti na iteraci. První iterace byla zvolena jako (20).

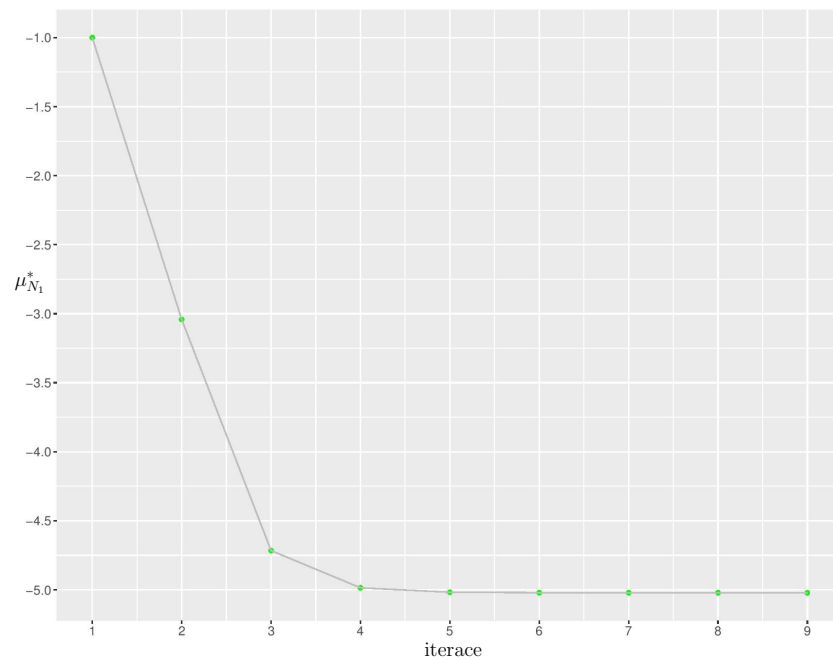
Z tabulky 1 a obrázku 8 lze vyčíst, že střední hodnoty generovaných dat a střední hodnoty dat variačních parametrů se téměř neliší. Rozptyly $\sigma_{N_1}^*$, $\sigma_{N_2}^*$ (viz tabulka 1 a obrázky 11, 12) jsou velmi malé, což znamená, že střední hodnoty jsou velmi přesné.



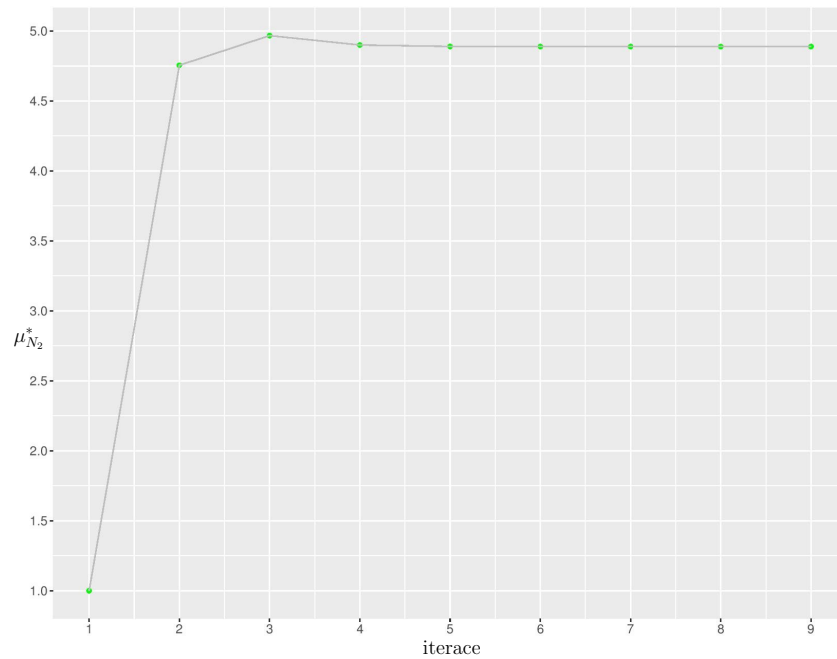
Obrázek 8: Histogram a hustota dat

Parametr	$\mu_{N_1}^*$	$\mu_{N_2}^*$	$\sigma_{N_1}^*$	$\sigma_{N_2}^*$	$d_{N_1}^*$	$d_{N_2}^*$
Hodnota	-5.021625	4.888685	0.07460192	0.02731082	54.57793	147.4221

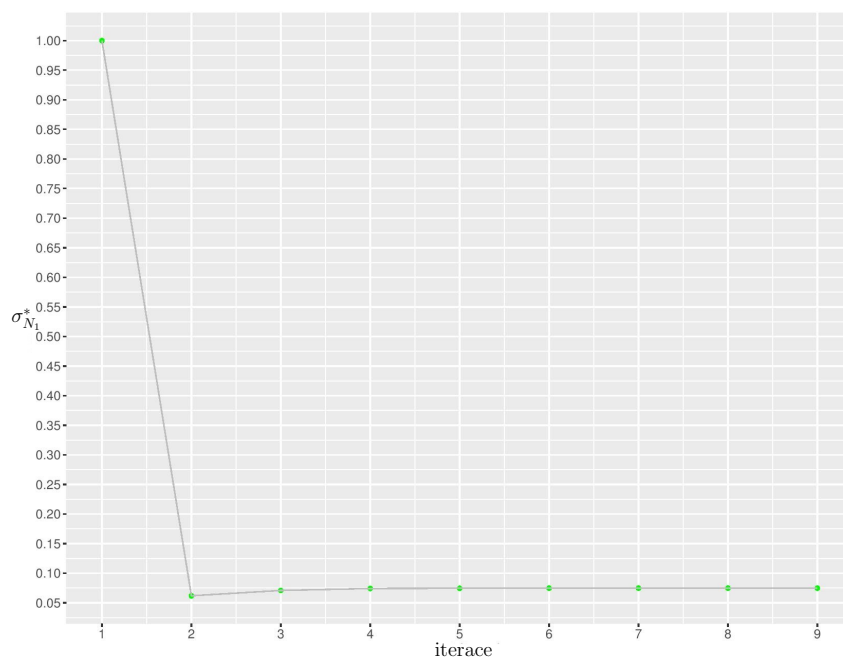
Tabulka 1: Tabulka variačních parametrů



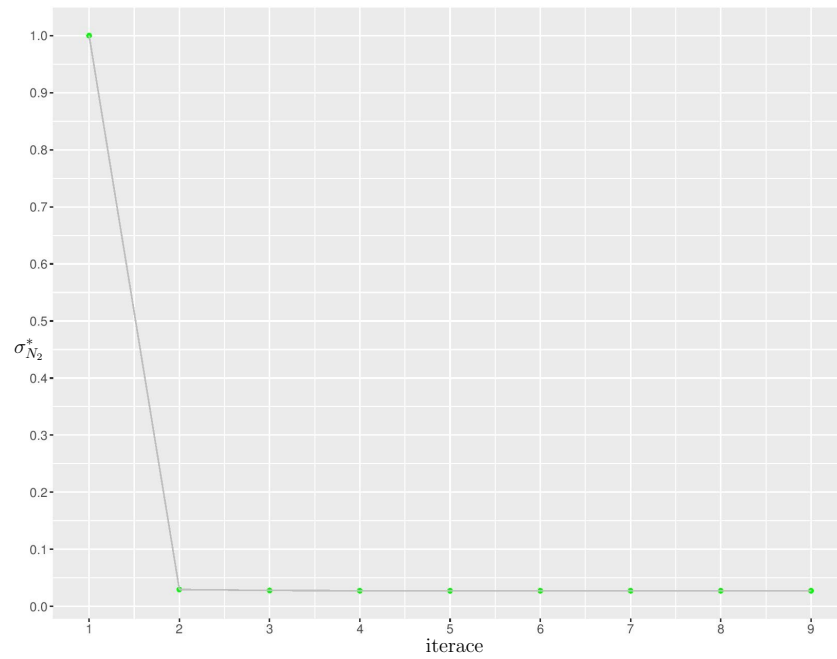
Obrázek 9: Odhad parametru $\mu_{N_1}^*$



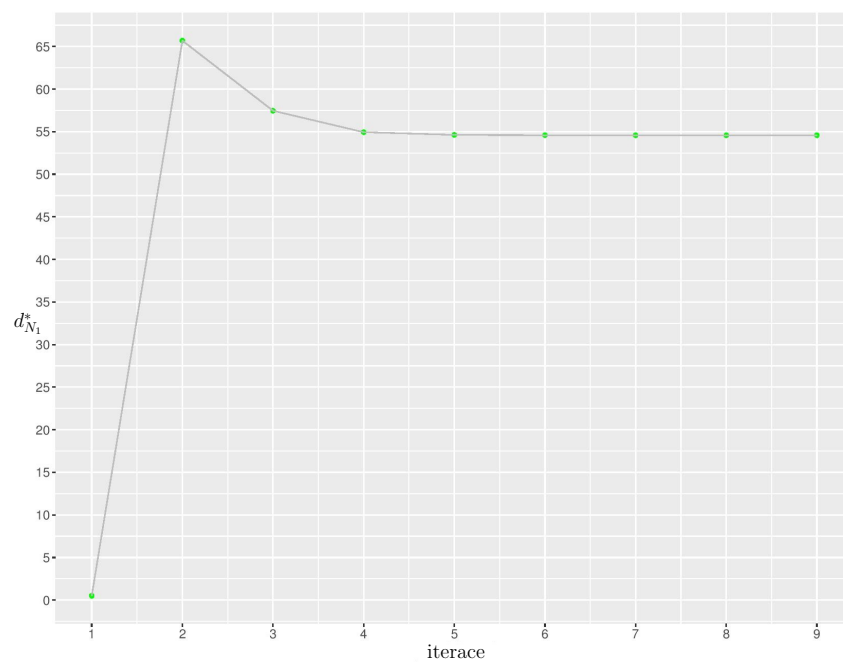
Obrázek 10: Odhad parametru $\mu_{N_2}^*$



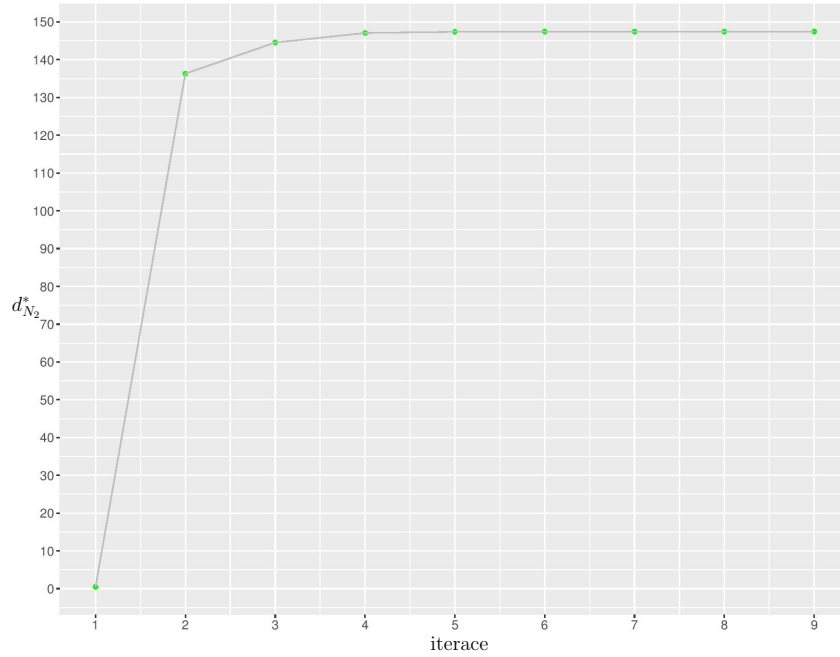
Obrázek 11: Odhad parametru $\sigma_{N_1}^*$



Obrázek 12: Odhad parametru $\sigma_{N_2}^*$



Obrázek 13: Odhad parametru $d_{N_1}^*$



Obrázek 14: Odhad parametru $d_{N_2}^*$

■

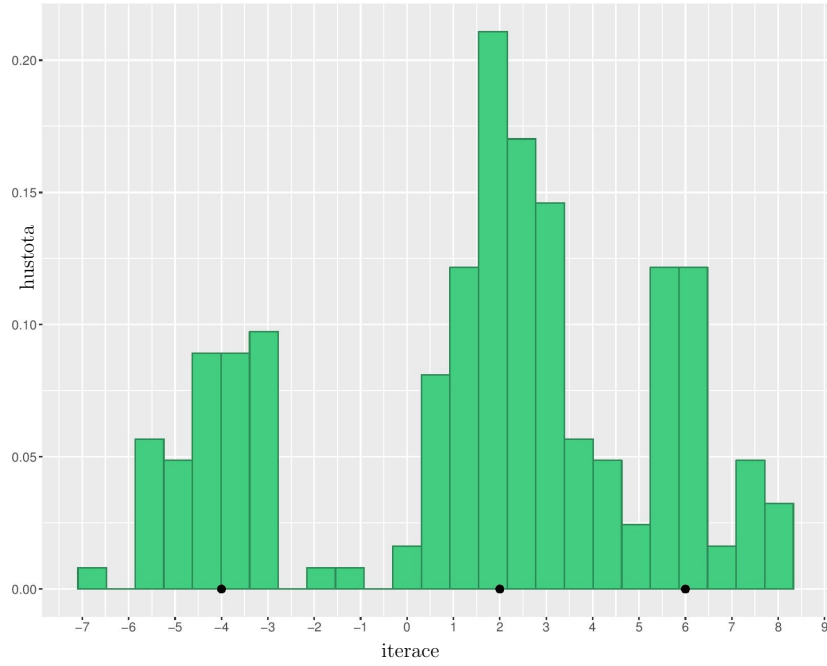
Příklad 4

Za pomocí algoritmu 2 budeme počítat variační parametry směsového modelu, který je formulován v kapitole 4.4. Použijeme simulovaná data ze směsi s 3 komponentami a počtem dat $N = 200$.

Střední hodnoty μ , přesnost τ a váhy α sloužící ke generování dat :

$$\begin{aligned}
 \tau &= 1 \\
 \alpha &= (0.3, 0.5, 0.2) \\
 \mu &= (-4, 2, 6)
 \end{aligned} \tag{21}$$

Na obrázku 15 lze vidět histogram dat nagenеровaných za pomocí těchto parametrů, společně s vyznačenými středními hodnotami pomocí černých puntíků.



Obrázek 15: Histogram nasimulovaných dat

Apriorní parametry byly zvoleny následovně:

$$\bar{\mu} = (-5, -3, 7)$$

$$k = (0.1, 0.1, 0.1)$$

$$d = (0.2, 0.3, 0.5)$$

$$a = 0.1$$

$$b = 0.1.$$

Jako počáteční aproximace variačních parametrů jsme zvolili následující hodnoty:

$$k_N^* = (1, 1, 1)$$

$$\mu_N^* = (-1, 0, 4)$$

$$d_N^* = (1, 1, 1) \tag{22}$$

$$a_N^* = 0.1$$

$$b_N^* = 0.1$$

Kritérium pro zastavení bylo zvoleno jako změna velikosti vektoru parametrů s prahem $\varepsilon = 10^{-4}$. Po 25 iteracích byly nalezeny variační parametry, které jsou uvedeny v tabulce 2. Na obrázku 16 je znázorněn histogram simulovaných dat (zeleně) generovaných z parametrů z

(21) a skutečná hustota pravděpodobnosti (zelená čára). Dále aproximace prediktivní hustoty

$$p(x|x_1, \dots, x_{200}) = \int p(x|\alpha_1, \alpha_2, \alpha_3, \mu_1, \mu_2, \mu_3, \tau) p(\alpha_1, \alpha_2, \alpha_3, \mu_1, \mu_2, \mu_3, \tau|x_1, \dots, x_{200}),$$

kde místo skutečné posteriorní hustoty $p(\alpha_1, \alpha_2, \alpha_3, \mu_1, \mu_2, \mu_3, \tau|x_1, \dots, x_{200})$ je použita nalezená variační aproximace, byla znázorněna oranžově. Dále jsou na obrázku puntíky znázorněny střední hodnoty, odlišené opět zelenou a oranžovou barvou. Jak lze vidět, data generována pomocí zvolených parametrů (21) jsou velmi podobná datům prediktivním.

Jak lze vidět na obrázku 16 a z tabulky 2, střední hodnoty simulovaných dat a střední hodnoty dat variačních parametrů jsou si velmi podobné.

Střední hodnota gamma rozdělení je rovna podílu variačních parametrů $\frac{a_N^*}{b_N^*}$.

Tedy $E_q[\tau] = \frac{a_N^*}{b_N^*} = \frac{100.1}{99.70982} = 1,003913$. Přesnost variačního rozdělení je tedy velmi blízká skutečné přesnosti.

Z těchto poznatků a obrázku 16 vidíme, že jsme došli ke správným výsledkům.

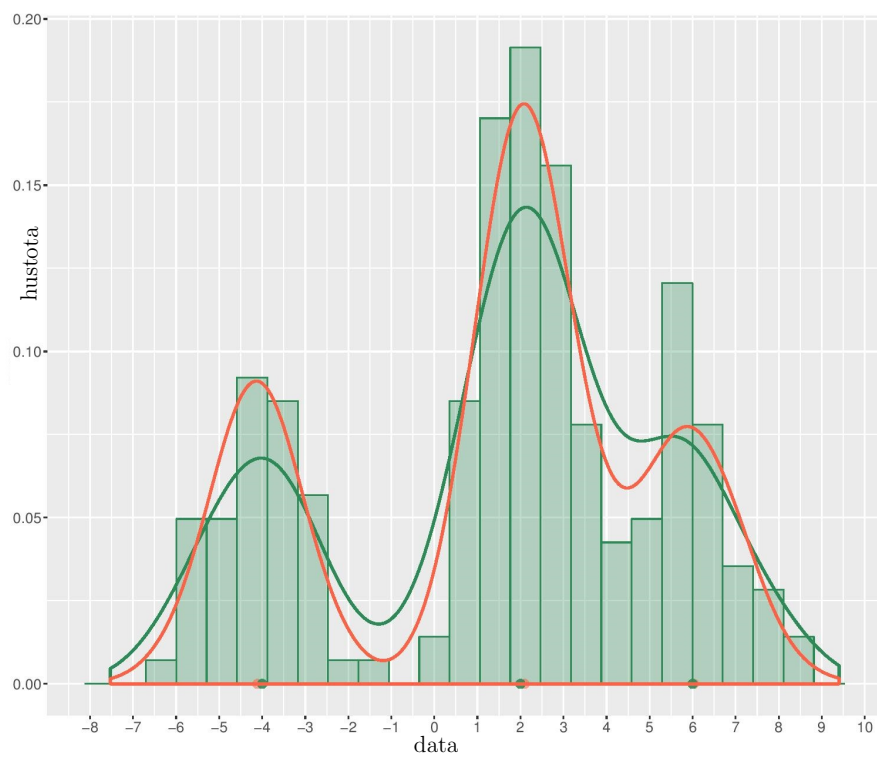
Na obrázcích 17 až 27 je znázorněn vývoj parametrů variační hustoty v závislosti na iteraci.

První iterace je dána počáteční volbou (22).

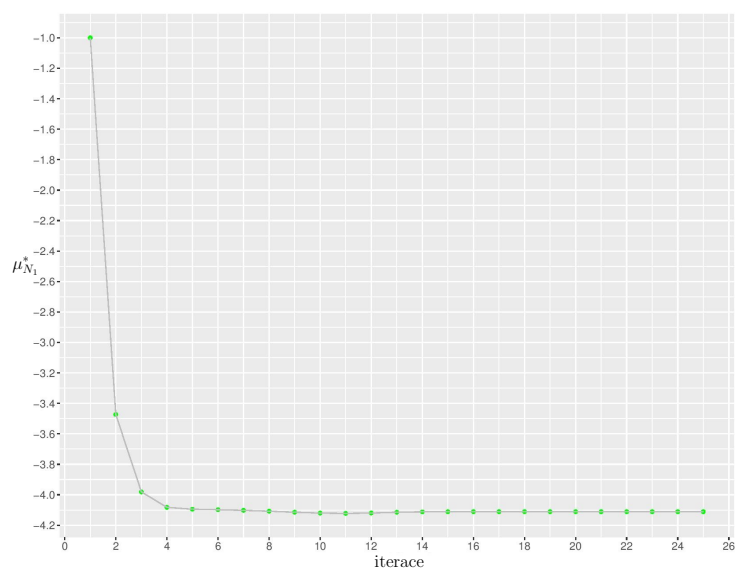
Parametr	$\mu_{N_1}^*$	$\mu_{N_2}^*$	$\mu_{N_3}^*$	$k_{N_1}^*$	$k_{N_2}^*$	$k_{N_3}^*$
Hodnota	-4.111024	2.093454	6.029711	49.99934	100.93651	49.36415

Parametr	a_N^*	b_N^*	$d_{N_1}^*$	$d_{N_2}^*$	$d_{N_3}^*$
Hodnota	100.1	99.70982	50.09934	101.13651	49.76415

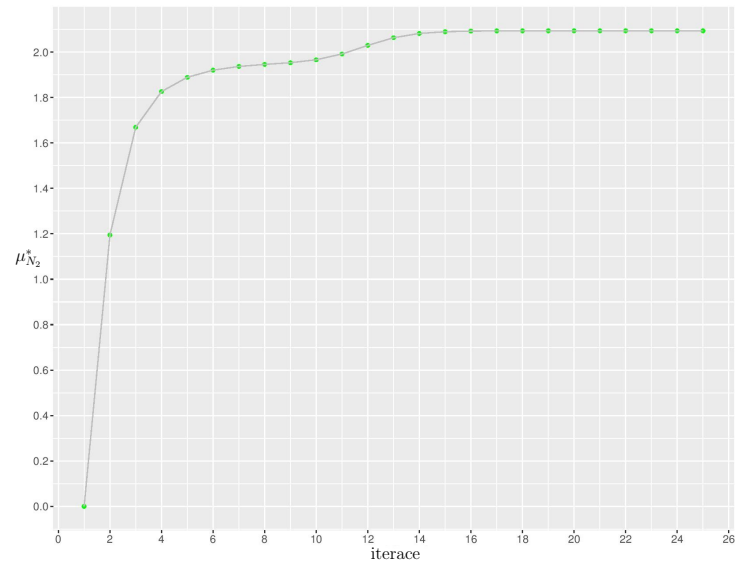
Tabulka 2: Tabulka variačních parametrů



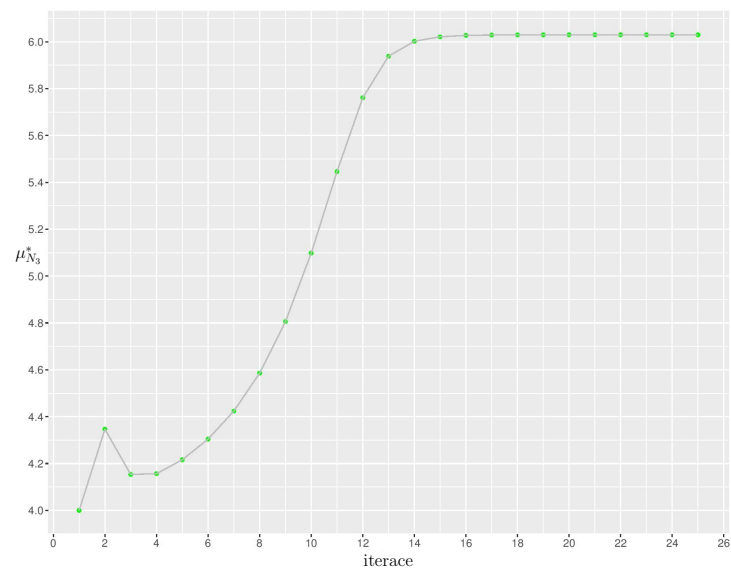
Obrázek 16: Histogram a hustoty dat



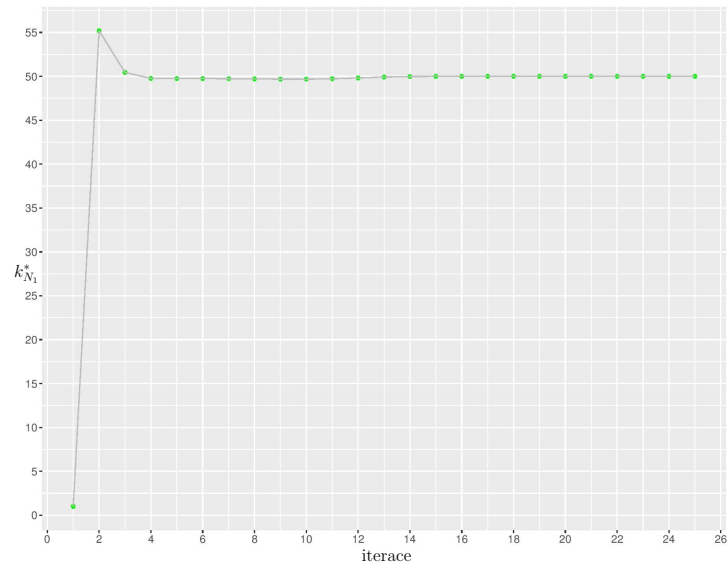
Obrázek 17: Odhad parametru $\mu_{N_1}^*$



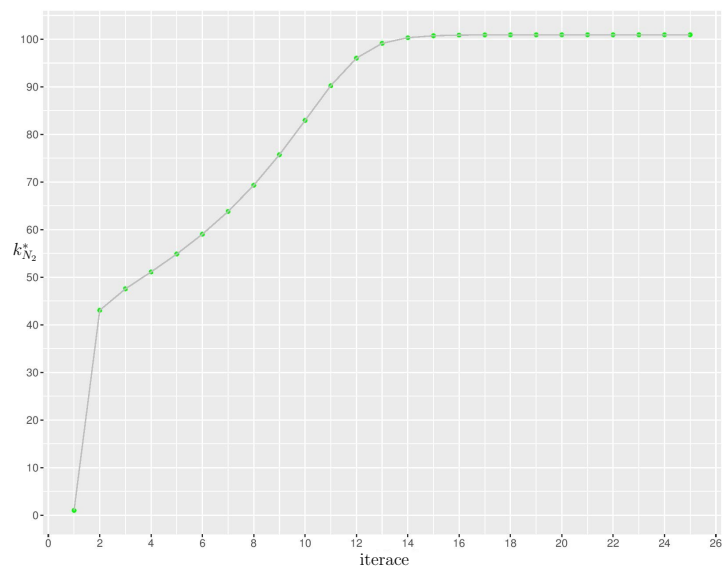
Obrázek 18: Odhad parametru $\mu_{N_2}^*$



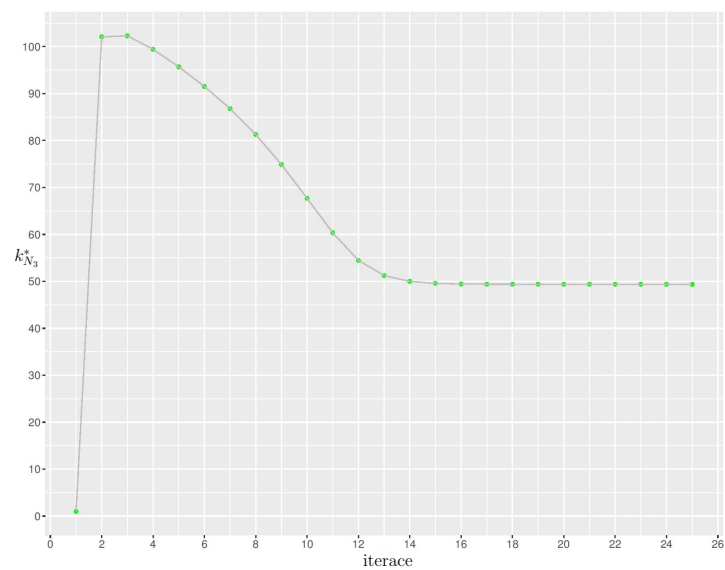
Obrázek 19: Odhad parametru $\mu_{N_3}^*$



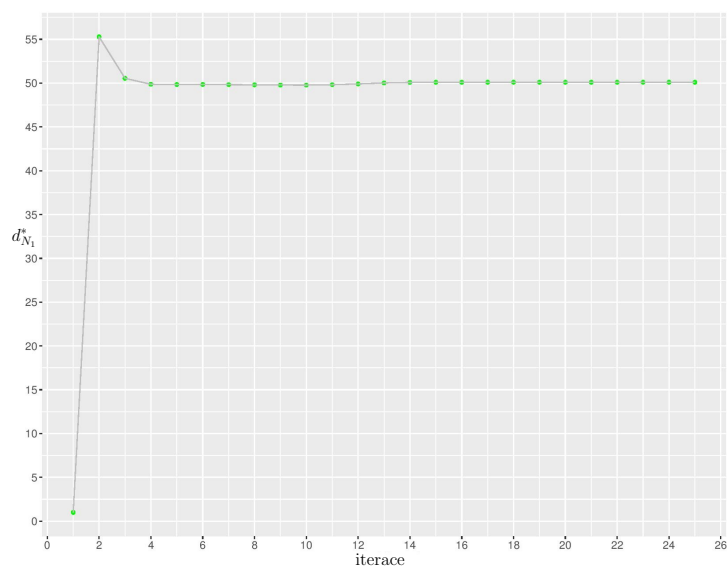
Obrázek 20: Odhad parametru $k_{N_1}^*$



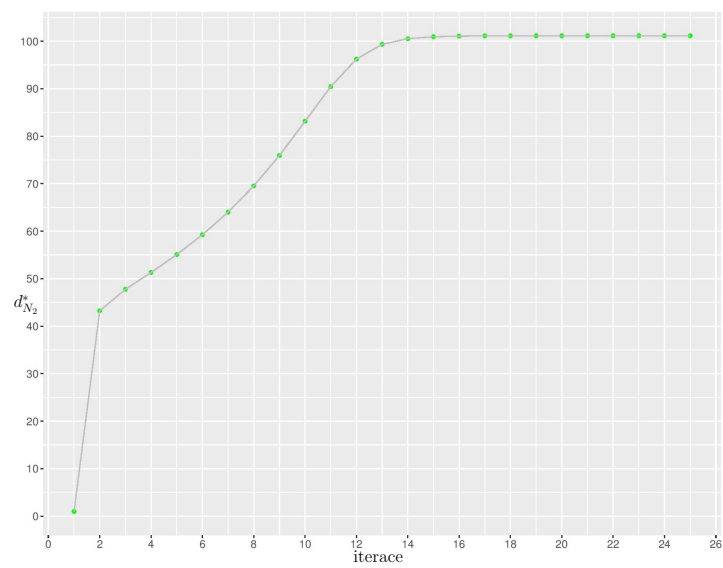
Obrázek 21: Odhad parametru $k_{N_2}^*$



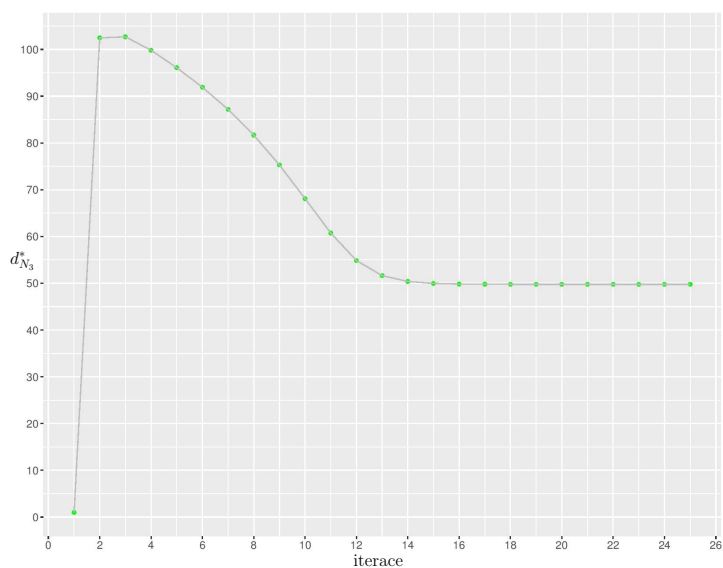
Obrázek 22: Odhad parametru $k_{N_3}^*$



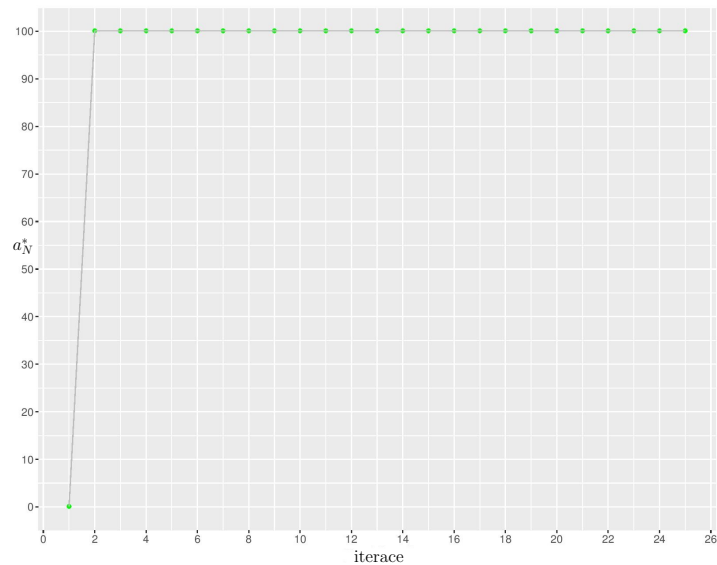
Obrázek 23: Odhad parametru $d_{N_1}^*$



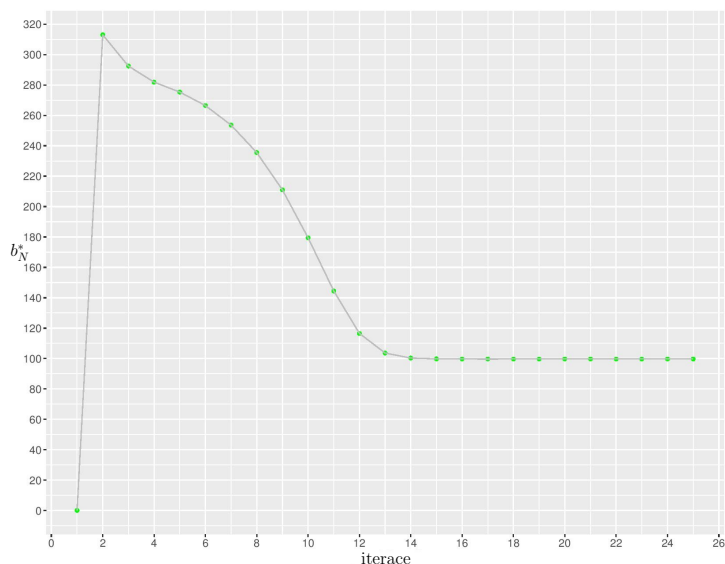
Obrázek 24: Odhad parametru $d_{N_2}^*$



Obrázek 25: Odhad parametru $d_{N_3}^*$



Obrázek 26: Odhad parametru a_N^*



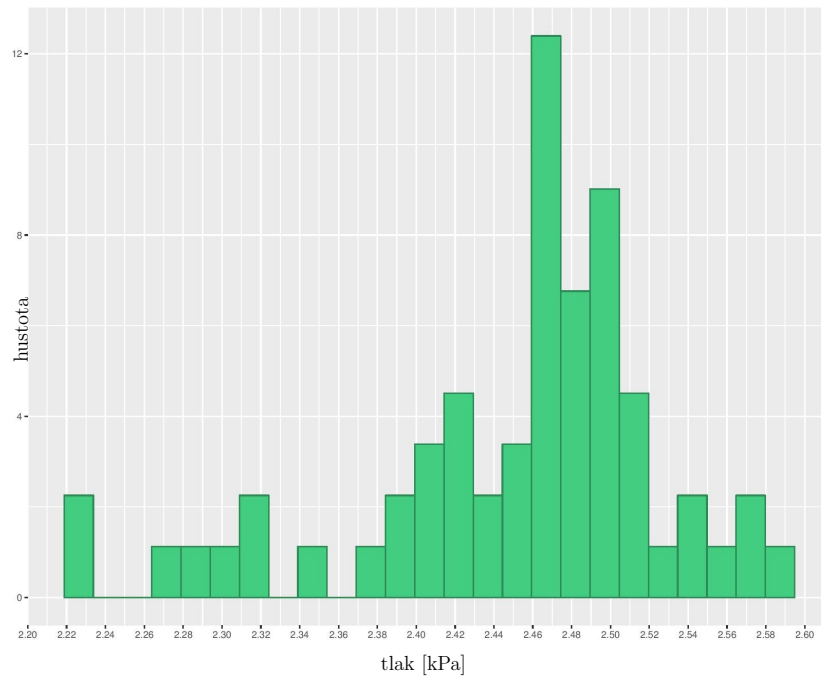
Obrázek 27: Odhad parametru b_N^*

■

Příklad 5

Nyní aplikujeme algoritmus 3 na reálná data, kterými jsou krevní tlaky pacienta, které byly měřeny v přibližně pravidelných intervalech v průběhu jednoho dne.

Měření bylo prováděno v rámci pravidelného vyšetření pacienta ve Fakultní nemocnici u Sv. Anny a anonymizovaná data byla dále využita pro výzkum se souhlasem pacienta dle postupu schváleného etickou komisí Fakultní nemocnice u Sv. Anny. Autorka práce pracovala s už anonymizovanými daty poskytnutými Dr. Polzerem.



Obrázek 28: Histogram tlakových dat

Data lze chápat jako 2 komponentní směs, kde do první komponenty patří tlaky měřené v noci a do druhé tlaky měřené přes den, přičemž denní tlaky mají větší váhu. Počet naměřených dat je $N = 59$. Jedná se opět o směsový model popsany v kapitole 4.4.

Data jsou popsána histogramem na obrázku 28

Apriorní parametry jsme si zvolili následovně:

$$\bar{\mu} = (2.1, 2.5)$$

$$k = (1, 1)$$

$$d = (0.5, 0.5)$$

$$a = 1$$

$$b = 1/400.$$

Volba počáteční aproximace:

$$k_N^* = 1$$

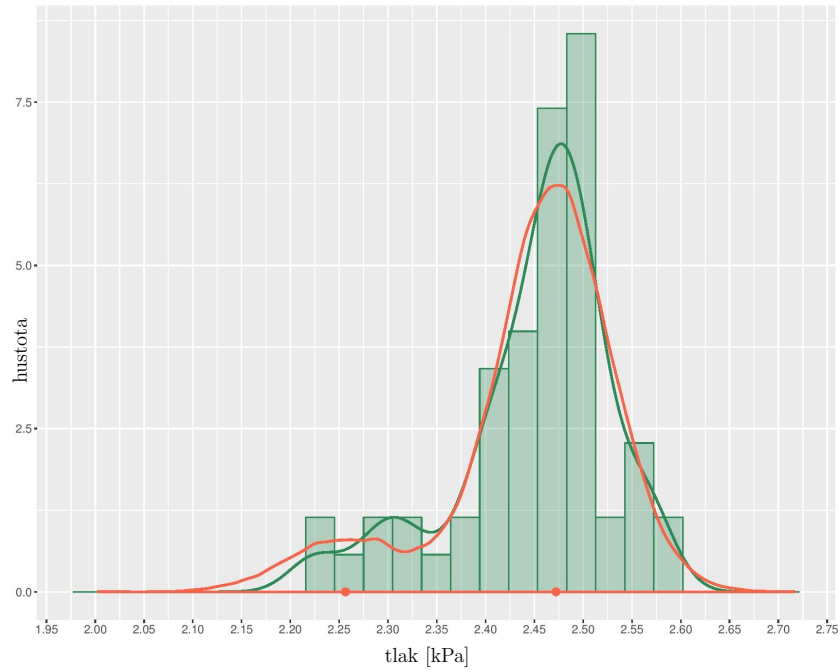
$$\mu_N^* = (2, 2.5)$$

$$d_N^* = 1$$

$$a_N^* = 1$$

$$b_N^* = 1/400$$

Pro zastavení výpočtu jsme opět použili změnu velikosti vektoru parametrů s prahem $\varepsilon = 10^{-4}$.

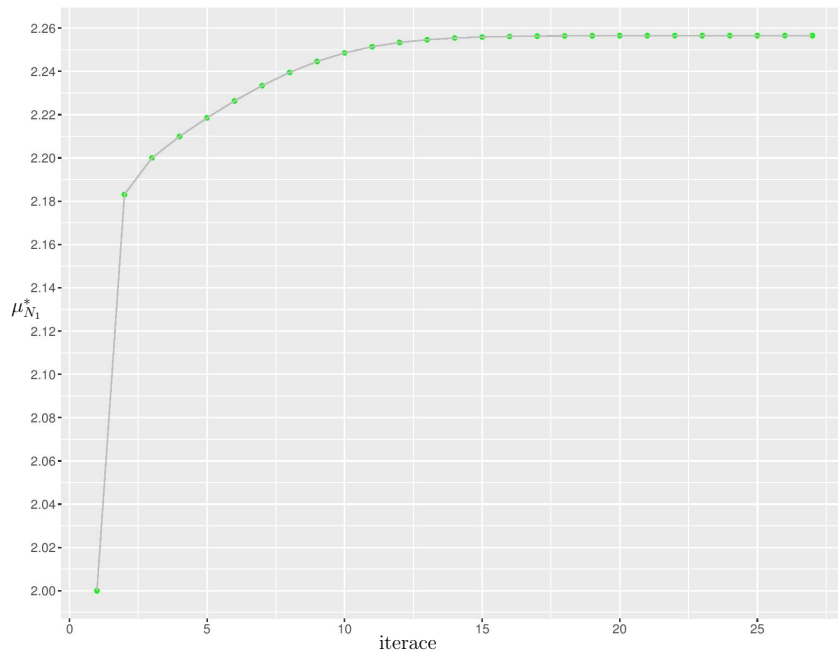


Obrázek 29: Histogram a hustoty tlakových dat

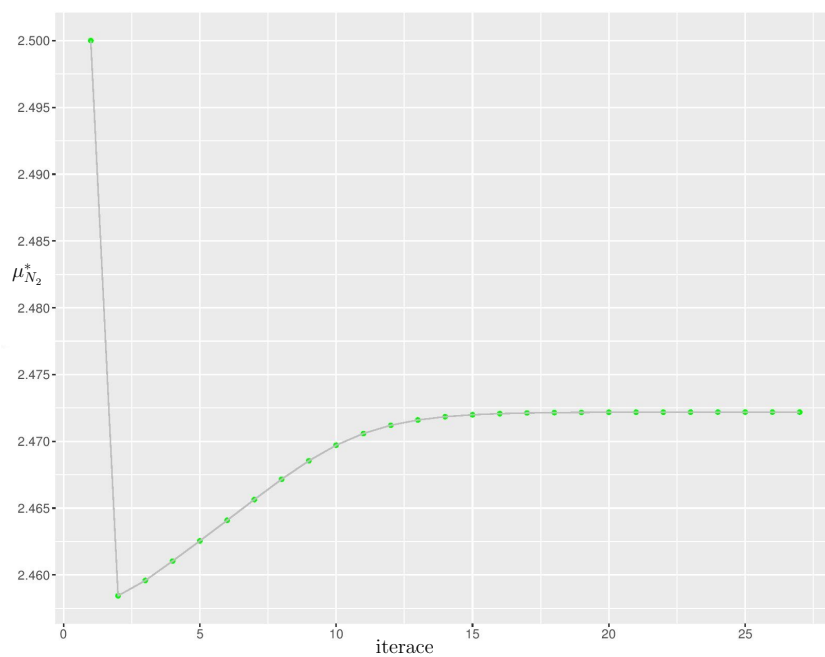
Variační parametry, uvedené v tabulce3 , byly nalezeny po 27 iteracích. Obrázek 29 ukazuje histogram tlakových dat (zeleně), hustotu tlakových dat(zeleně) a aproximaci prediktivní hustoty (oranžově). Střední hodnoty prediktivních dat jsou znázorněny oranžovými puntíky. Na obrázcích 30 až 36 můžeme vidět změnu parametrů variační hustoty v závislosti na iteraci. První iterace odpovídá počáteční volbě 5.

Parametr	$\mu_{N_1}^*$	$\mu_{N_2}^*$	$k_{N_1}^*$	$k_{N_2}^*$	a_N^*	b_N^*	$d_{N_1}^*$	$d_{N_2}^*$
Hodnota	-2.256513	2.472186	7.70831	53.29169	7.20831	52.79169	30.5	0.09303221

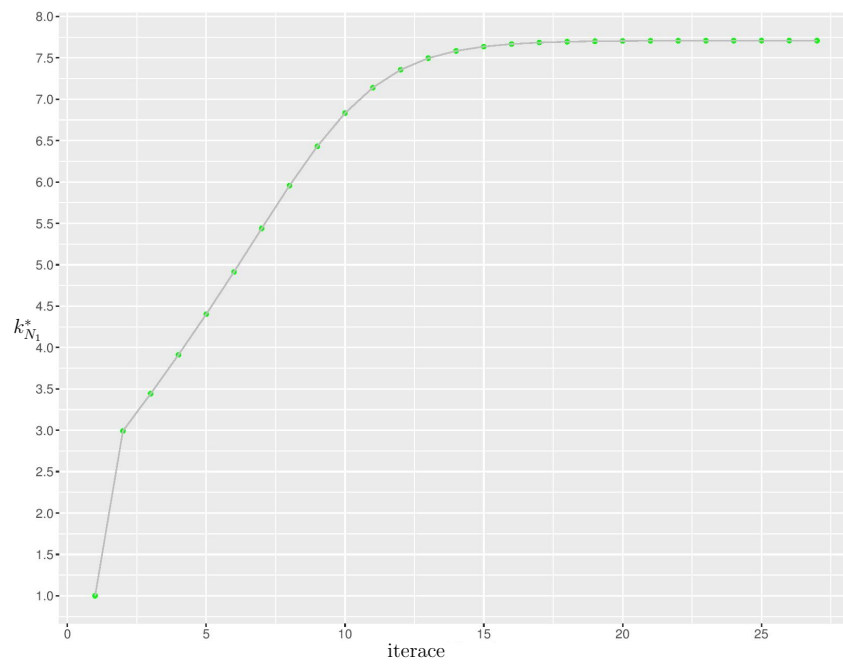
Tabulka 3: Tabulka variačních parametrů tlakových dat



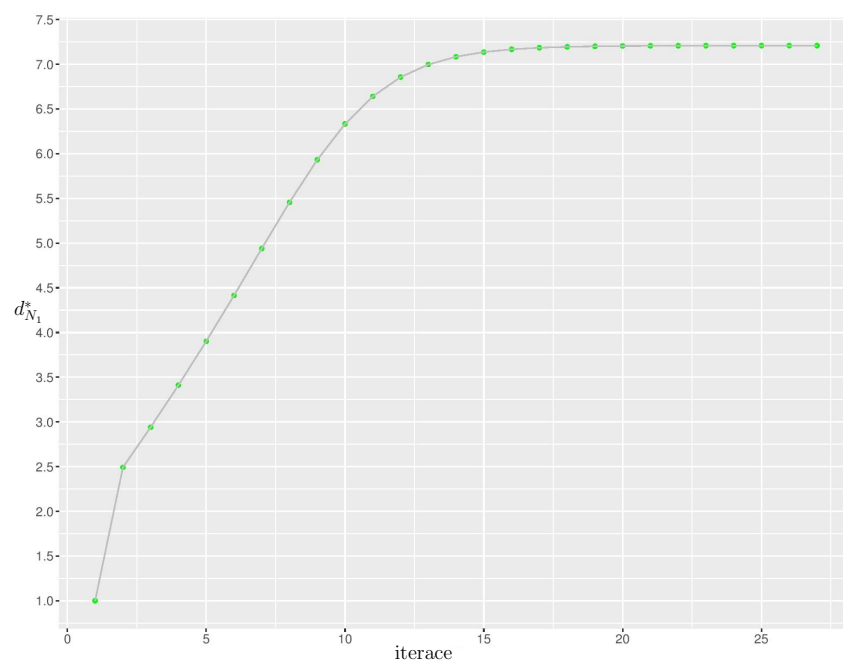
Obrázek 30: Odhad parametru $\mu_{N_1}^*$



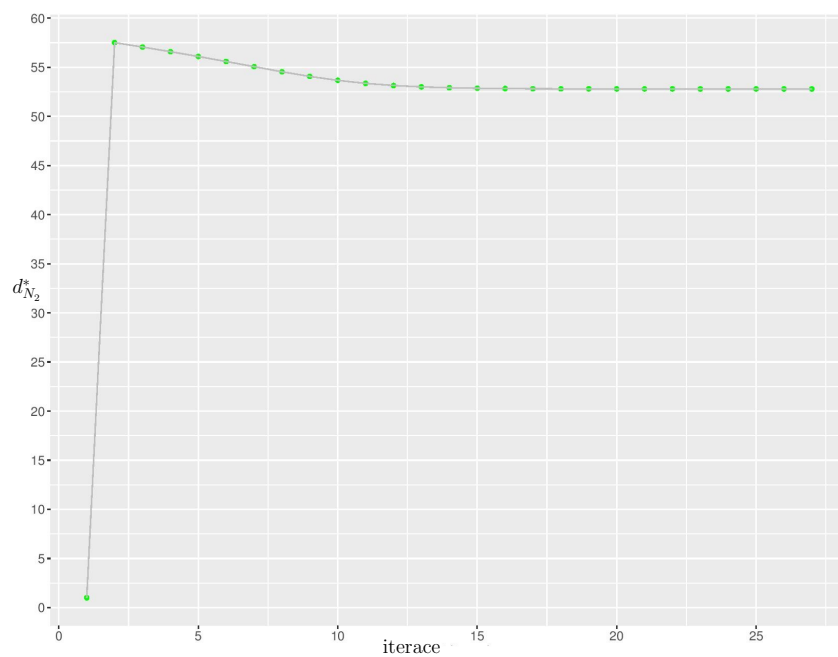
Obrázek 31: Odhad parametru $\mu_{N_2}^*$



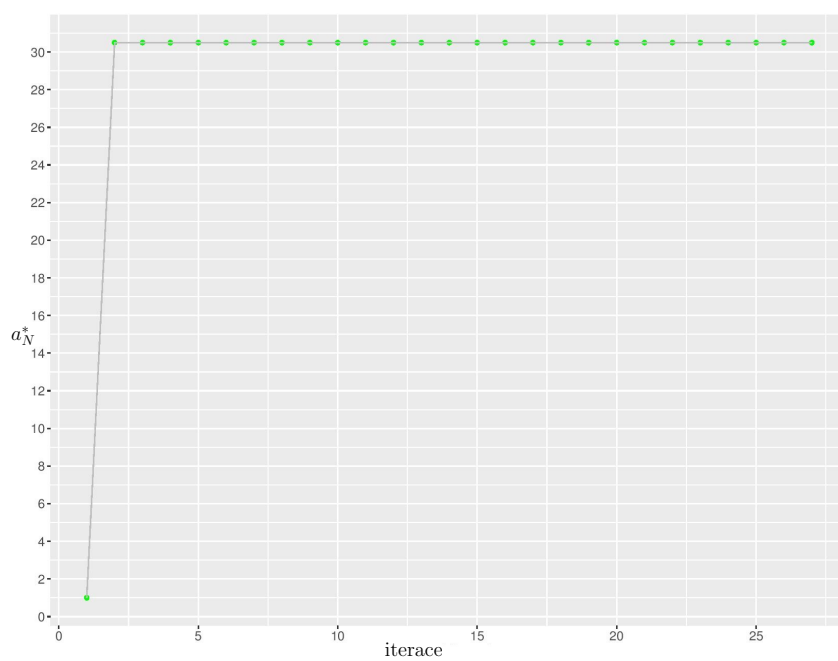
Obrázek 32: Odhad parametru $k_{N_1}^*$



Obrázek 33: Odhad parametru $d_{N_1}^*$

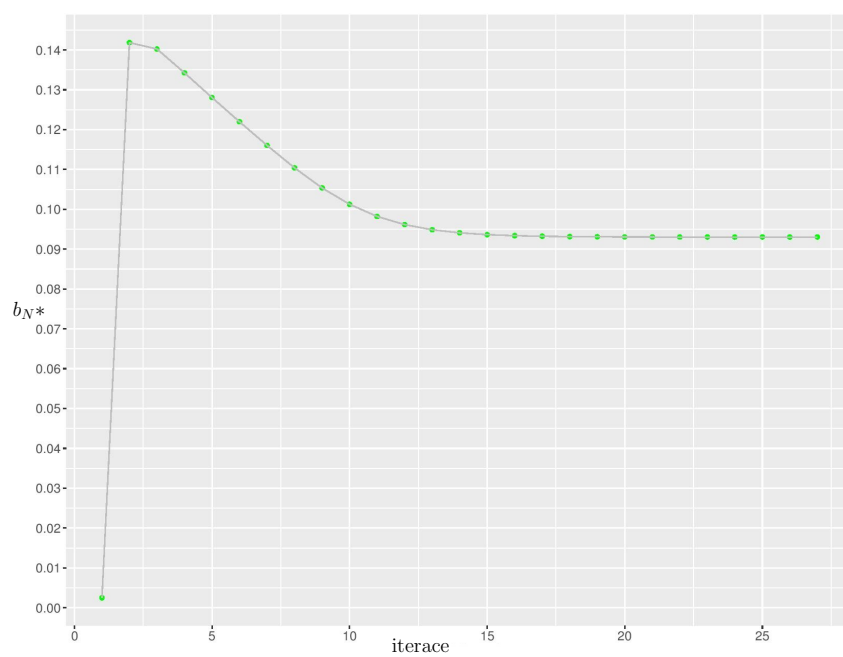


Obrázek 34: Odhad parametru $d_{N_2}^*$



Obrázek 35: Odhad parametru a_N^*

■



Obrázek 36: Odhad parametru b_N^*

6 Závěr

Cílem této práce bylo seznámit se s principy variačních bayesovských metod a následně je aplikovat na směšové modely. Součástí práce byla implementace algoritmů, které byly využity pro nalezení variačních parametrů daných směšových modelů. Algoritmus jsme aplikovali nejen na simulovaná data, ale i na data reálná. I přes velkou náročnost odvození variační aproximace jsme došli ke správným výsledkům, což lze vyčíst z uvedených. Algoritmy našly hledané variační parametry po velmi malém počtu iterací (≈ 20 iterací). V porovnání s MCMC, kde je potřeba generovat tisíce, někdy až miliony vzorků, je variační aproximace nesrovnatelně rychlejší.

Literatura

- [1] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.
- [2] Charles W Fox and Stephen J Roberts. A tutorial on variational bayesian inference. *Artificial intelligence review*, 38(2):85–95, 2012.
- [3] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.
- [4] Kevin P Murphy. Conjugate bayesian analysis of the gaussian distribution. *def*, 1(2 σ 2):16, 2007.
- [5] Thomas Dyhre Nielsen and Finn Verner Jensen. *Bayesian networks and decision graphs*. Springer Science & Business Media, 2009.
- [6] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [7] Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.
- [8] Josef Štěpán. *Teorie pravděpodobnosti: Matematické základy*. Academia, 1987.
- [9] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

A Popis algoritmů

Algoritmy 1 (algoritmus_2_komponentni_smes.R), 2 (algoritmus_n_komponentni_smes.R) a 3 (algoritmus_tlaky.R) jsou na přiloženém CD. Pro funkčnost algoritmu 3 je nutné načíst přiložený csv soubor s názvem tlaky.csv.

```
//Algoritmus pocitajici variacni parametry pro 2 komponentni smes
library(ggplot2)
library(MCMCpack)
set.seed(10)

//pocet dat
N<-200

//apriorni data
mu1<- -3
mu2<- 3
sigma1<-10
sigma2<-10
d1<-1
d2<-1

//parametry pro simulaci dat
mu1<- -5
mu2<- 5
sigma1<-2
sigma2<-2
sd<-c(sigma1, sigma2)
alpha1<-0.3
mu<- c(mu1,mu2)
sd_<- c(sigma1_,sigma2_)

//prah
eps<-0.0001

//generovani dat
components<-sample(1:2,prob=c(alpha1,1-alpha1),size=N,replace=TRUE)
x<-rnorm(n=N,mean=mu[components],sd[components])
```

```

// parametry variacni aproximace
mu_N1.<- vector()
mu_N2.<- vector()
sigma_N1.<- vector()
sigma_N2.<- vector()
d_N1.<- vector()
d_N2.<- vector()
theta1<- vector()
theta2<- vector()
phi1<- vector()
phi2<- vector()
lambda1<- vector()
lambda2<- vector()
lambda3<- vector()
lambda4<- vector()
lambda5<- vector()
lambda6<- vector()

// parametry vychozi aproximace
mu_N1.[1]<-(-1)
mu_N2.[1]<-1
sigma_N1.[1]<-1
sigma_N2.[1]<-1
d_N1.[1]<-0.5
d_N2.[1]<-0.5
i<-2

// parametry variacni aproximace v druhe iteraci
mu_N1.[2]<-mu_N1.[1]+1
mu_N2.[2]<-mu_N2.[1]+1
sigma_N1.[2]<-sigma_N1.[1]+1
sigma_N2.[2]<-sigma_N2.[1]+1
d_N1.[2]<-d_N1.[1]+1
d_N2.[2]<-d_N2.[1]+1

// vektor normy rozdilu variacniho parametru v i iteraci a i-1 iteraci
rozdil_mu_N1.<- vector()
rozdil_mu_N2.<- vector()
rozdil_sigma_N1.<- vector()

```

```

rozdil_sigma_N2.<- vector()
rozdil_d_N1.<- vector()
rozdil_d_N1.<- vector()

rozdil_mu_N1.<-sqrt((((sqrt((mu_N1.[i])^2))-(sqrt((mu_N1.[i-1])^2))))^2)
rozdil_mu_N2.<-sqrt((((sqrt((mu_N2.[i])^2))-(sqrt((mu_N2.[i-1])^2))))^2)
rozdil_sigma_N1.<-sqrt((((sqrt((sigma_N1.[i])^2))-(sqrt((sigma_N1.[i-1])^2))))^2)
rozdil_sigma_N2.<-sqrt((((sqrt((sigma_N2.[i])^2))-(sqrt((sigma_N2.[i-1])^2))))^2)
rozdil_d_N1.<-sqrt((((sqrt((d_N1.[i])^2))-(sqrt((d_N1.[i-1])^2))))^2)
rozdil_d_N2.<-sqrt((((sqrt((d_N2.[i])^2))-(sqrt((d_N2.[i-1])^2))))^2)

while ((rozdil_mu_N1. > eps)|(rozdil_mu_N2. > eps)|(rozdil_sigma_N1. > eps)|
(rozdil_sigma_N2. > eps)|(rozdil_d_N1. > eps)|(rozdil_d_N1. > eps)) {
for (n in 1:N) {
//vypocet lokálnych variacných parametrov

phi1[n]<-(digamma(d_N1.[i-1])-digamma(d_N1.[i-1]+d_N2.[i-1]))
-log(sqrt(2*pi*(sigma1^2)))+(((((-x[n])^2+2*x[n]*mu_N1.[i-1]
-(sigma_N1.[i-1])^2-(mu_N1.[i-1])^2))/(2*(sigma1^2)))

phi2[n]<-(digamma(d_N2.[i-1])-digamma(d_N1.[i-1]+d_N2.[i-1]))
-log(sqrt(2*pi*(sigma2^2)))+(((((-x[n])^2+2*x[n]*mu_N2.[i-1]
-(sigma_N2.[i-1])^2-(mu_N2.[i-1])^2))/(2*(sigma2^2)))

theta1[n]<-exp(phi1[n])
theta2[n]<-exp(phi2[n])

}
//normalizace parametrov theta
theta12=theta1+theta2
theta1=theta1/theta12
theta2=theta2/theta12

//vypocet globalných variacných parametrov
lambda1[i]<-(mu1/(sigma1^2))+(sum(theta1[1:N]*x[1:N]))/(sigma1^2)
lambda2[i]<-(-1/2)*((1/sigma1^2)+(sum(theta1[1:N]))/(sigma1^2))
mu_N1.[i]<-(-lambda1[i])/(2*lambda2[i])
sigma_N1.[i]<-(-1/(2*lambda2[i]))

```

```

lambda3[i]<-(mu2_/(sigma2_^2))+(sum(theta2[1:N]*x[1:N]))/(sigma2^2)
lambda4[i]<-(-1/2)*((1/sigma2_^2)+(sum(theta2[1:N]))/(sigma2^2))
mu_N2.[i]<-(-lambda3[i])/(2*lambda4[i])
sigma_N2.[i]<-(-1/(2*lambda4[i]))

lambda5[i]<-d1+sum(theta1[1:N])
lambda6[i]<-d2+sum(theta2[1:N])
d_N1.[i]<-lambda5[i]
d_N2.[i]<-lambda6[i]

// norma rozdilu variacniho parametru v i iteraci a i-1 iteraci
rozdil_mu_N1.<-sqrt(((sqrt((mu_N1.[i])^2))-sqrt((mu_N1.[i-1])^2)))^2)
rozdil_mu_N2.<-sqrt(((sqrt((mu_N2.[i])^2))-sqrt((mu_N2.[i-1])^2)))^2)
rozdil_sigma_N1.<-sqrt(((sqrt((sigma_N1.[i])^2))-sqrt((sigma_N1.[i-1])^2)))^2)
rozdil_sigma_N2.<-sqrt(((sqrt((sigma_N2.[i])^2))-sqrt((sigma_N2.[i-1])^2)))^2)
rozdil_d_N1.<-sqrt(((sqrt((d_N1.[i])^2))-sqrt((d_N1.[i-1])^2)))^2)
rozdil_d_N2.<-sqrt(((sqrt((d_N2.[i])^2))-sqrt((d_N2.[i-1])^2)))^2)
i<-i+1

}

//aktualni maximalni iterace
n_it=i-1

//znazorneni variacnich parametru v dane iteraci
ggplot(data=NULL,aes(x =j<-1:(i-1) , y = mu_N1.)) + geom_point(color = 'green')
+ geom_line(color="grey")+labs(x="pocet iteraci")
+ scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+ scale_y_continuous(breaks = scales::pretty_breaks(n =15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x =j<-1:(i-1) , y = mu_N2.)) + geom_point(color = 'green')
+ geom_line(color="grey")+labs(x="pocet iteraci")
+ scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = j<-1:(i-1), y = sigma_N1.))
+geom_point(color = 'green') + geom_line(color="grey")+labs(x="pocet iteraci")

```

```

+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = j<-1:(i-1), y = sigma_N2.))+geom_point(color = 'green'
)
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = j<-1:(i-1), y = d_N1.))+geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = j<-1:(i-1), y = d_N2.))
+geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

//znazorneni histogramu nasimulovanych dat
ggplot(data=NULL,aes(x,y=..density..))
+geom_histogram(color="seagreen", fill="seagreen3",bins = 25)
+geom_point(data=NULL,aes(x=mu, y=0), colour="black",size=3)
+scale_x_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

//generovani dat s variacnimi parametry
x_<-vector()
sd_n = vector()
for(s in 1:n_it)
{
sd_n=c(sigma_N1.[n_it],sigma_N2.[n_it])

```

```

mu_n=c(mu_N1.[n_it],mu_N2.[n_it])
d_n=c(d_N1.[n_it],d_N2.[n_it])
alpha_=rdirichlet(1, d_n)
muP=rnorm(2, mu_n, sqrt(sd_n))

components_<-sample(1:2,prob=alpha_,size=N,replace=TRUE)
x_<-c(x_, rnorm(n=N,mean=muP[components_],sd))
}

p_1=ggplot()

//znazorneni histogramu pro data se skutecnymi parametry
p_2=geom_histogram(data=NULL,aes(x,y=..density..),alpha=0.3,color="seagreen4",
  fill="seagreen",bins = 25)

//znazorneni strednich hodnot pro data se skutecnymi parametry
p_4=geom_point(data=NULL,aes(x=mu, y=0),alpha=0.9, colour="seagreen",size=3)

//znazorneni strednich hodnot pro data s variacnimi parametry
p_5=geom_point(data=NULL,aes(x=mu_n, y=0),alpha=0.7, colour="tomato",size=3)

//znazorneni hustoty pro data se skutecnymi parametry
p_6=geom_density(data=NULL,aes(x),alpha=0.2,color="seagreen",size=1)

//znazorneni hustoty pro data s variacnimi parametry
p_7=geom_density(data=NULL,aes(x_),alpha=0.2,color="tomato",size=1)

p_8=scale_x_continuous(breaks = scales::pretty_breaks(n = 15))
p_9=theme(axis.text=element_text(size=11),axis.title=element_text(size=14))

p_1+p_2+p_4+p_5+p_6+p_7+p_8+p_9+labs(x = "data",y="hustota")

//vypsani variacnich parametru v posledni iteraci
print(mu_N1.[n_it])
print(mu_N2.[n_it])
print(sigma_N1.[n_it])
print(sigma_N2.[n_it])
print(d_N1.[n_it])
print(d_N2.[n_it])

```



```
print(n_it)
```

Výpis 1: Algoritmus počítající variační parametry pro 2 komponentní směs

```
//Algoritmus pocitajici variacni parametry pro n komponentni smes
```

```
library(ggplot2)
```

```
library(MCMCpack)
```

```
set.seed(5)
```

```
//pocet dat
```

```
N=200
```

```
//pocet komponent
```

```
J=3
```

```
//maximalni pocet iteraci
```

```
n_it=1000
```

```
//prah
```

```
eps=0.00001
```

```
// apriorni parametry
```

```
mu_=c(-5,-3, 7)
```

```
k=c(0.1,0.1,0.1)
```

```
d=c(0.2,0.3,0.5)
```

```
a=0.1
```

```
b=0.1
```

```
// parametry pro simulaci dat
```

```
tau<-1
```

```
alpha<-c(0.3,0.5,0.2)
```

```
sd <- vector()
```

```
for(l in 1:J) {sd[l] <- 1/sqrt(tau)}
```

```
mu<-c(-4,2, 6)
```

```
// generovani dat
```

```
components<-sample(1:J,prob=alpha,size=N,replace=TRUE)
```

```
x<-rnorm(n=N,mean=mu[components],sd[components])
```

```

// parametry variacni aproximace
lambda<-matrix(nrow=n_it,ncol=2+3*J)
phi<-matrix(nrow=N,ncol=J)

theta<-matrix(nrow=N,ncol=J)

k_N.<-matrix(nrow=n_it,ncol=J)
mu_N.<-matrix(nrow=n_it,ncol=J)
d_N.<-matrix(nrow=n_it,ncol=J)
a_N.<-vector()
b_N.<-vector()

// parametry vychozi aproximace
k_N.[1,1:J]<-1
mu_N.[1,1:J]<-c(-1,0,4)
d_N.[1,1:J]<-1
a_N.[1]<-0.1
b_N.[1]<-0.1

i<-2

rozdil_mu_N.<-vector()
rozdil_d_N.<-vector()
rozdil_k_N.<-vector()

podminka=0
while(podminka==0)
{

for(j in 1:J)
{
for(n in 1:N)
{
//vypocet lokalnich variacnich parametru
phi[n,j]=digamma(d_N.[i-1,j])-digamma(sum(d_N.[i-1,1:J]))
-log(sqrt(2*pi))+(1/2)*(digamma(a_N.[i-1])-log(b_N.[i-1]))
+((-a_N.[i-1]/(2*b_N.[i-1]))*x[n]^2)+(a_N.[i-1]/b_N.[i-1])*mu_N.[i-1,j]*x[n]
-(1/2)*(1/k_N.[i-1,j]+(a_N.[i-1]/b_N.[i-1])*mu_N.[i-1,j]^2))

```

```

theta[n,j]=exp(phi[n,j])
}
}
//normalizace parametru theta
theta=theta/rowSums(theta)

//vypocet globalnich variacnich parametru
for(j in 1:J)
{
lambda[i,2+j]=k[j]*mu_[j]+sum(x[1:N]*theta[1:N,j])
lambda[i,2+j+J]=(-1/2)*(k[j]+sum(theta[1:N,j]))
lambda[i,2+j+2*J]=d[j]+sum(theta[1:N,j])

k_N.[i,j]=-2*lambda[i,2+j+J]
mu_N.[i,j]=- (lambda[i,2+j])/(2*lambda[i,2+j+J])
d_N.[i,j]=lambda[i,2+j+2*J]

}

lambda[i,1]=a-1+J/2+N/2
lambda[i,2]=-b-1/2*(sum(x*x)+sum(mu_^2*k))

a_N.[i]=lambda[i,1]+1-J/2
b_N.[i]=(1/4)*sum(lambda[i,3:(2+J)]^2/lambda[i,(3+J):(2+2*J)])-lambda[i,2]

//ukoncujici kriterium - chyba normy rozdilu variacnich parametru v i a i-1
iteraci
podminka=1
for(j in 1:J)
{
rozdil_mu_N.[j]<-sqrt(((sqrt((mu_N.[i,j])^2))-sqrt((mu_N.[i-1,j])^2)))^2)
rozdil_d_N.[j]<-sqrt(((sqrt((d_N.[i,j])^2))-sqrt((d_N.[i-1,j])^2)))^2)
rozdil_k_N.[j]<-sqrt(((sqrt((k_N.[i,j])^2))-sqrt((k_N.[i-1,j])^2)))^2)

rozdil_b_N.<-sqrt(((sqrt((k_N.[i])^2))-sqrt((k_N.[i-1])^2)))^2)

if((rozdil_mu_N.[j])>eps || (rozdil_d_N.[j])>eps || (rozdil_k_N.[j])>eps ||
(rozdil_b_N.)>eps)

```

```

{
podminka=0
break
}
}

i<-i+1
}

//aktualni maximalni iterace
n_it<- i-1

//znazorneni variacnich parametru v dane iteraci
for (j_it in 1:J)
{
p1=ggplot(data=NULL,aes(x =m<-1:(n_it) , y = mu_N.[(1:n_it),j_it]))
+geom_point(color = 'green') + geom_line(color="grey")
+labs(x="pocet iteraci")+labs(y=paste("mu_N.",j_it))
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

p2=ggplot(data=NULL,aes(x = m<-1:(n_it), y = d_N.[(1:n_it),j_it]))
+geom_point(color = 'green') + geom_line(color="grey")
+labs(x="pocet iteraci")+ labs(y=paste("d_N.",j_it))
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

p3=ggplot(data=NULL,aes(x = m<-1:(n_it), y = k_N.[(1:n_it),j_it]))
+geom_point(color = 'green') + geom_line(color="grey")
+labs(x="pocet iteraci")+labs(y=paste("k_N.",j_it))
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

show(p1)

```

```

show(p2)
show(p3)
}

ggplot(data=NULL,aes(x = m<-1:(n_it), y = a_N.)) + geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = m<-1:(n_it), y = b_N.)) + geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

//znazorneni histogramu nasimulovanych dat
ggplot(data=NULL,aes(x,y=..density..))
+geom_histogram(color="seagreen", fill="seagreen3",bins = 25)
+geom_point(data=NULL,aes(x=mu, y=0), colour="black",size=3)
+scale_x_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

//generovani dat s variacnimi parametry
x_<-vector()
sd_ = vector()
for(s in 1:n_it)
{
tau_=rgamma(1, shape=a_N.[n_it],rate=b_N.[n_it])
for(j in 1:J)
{
sd_[j] <- 1/sqrt(k_N.[n_it,j]*tau_)
}
alpha_=rdirichlet(1, d_N.[n_it,])
muP=rnorm(J, mu_N.[n_it,], sd_)

components_<-sample(1:J,prob=alpha_,size=N,replace=TRUE)
x_<-c(x_, rnorm(n=N,mean=muP[components_],tau_**(-0.5)))

```

```

}

p_1=ggplot()

//znazorneni histogramu pro data se skutecnymi parametry
p_2=geom_histogram(data=NULL,aes(x,y=..density..),alpha=0.3,color="seagreen4",
  fill="seagreen",bins = 25)

//znazorneni strednich hodnot pro data se skutecnymi parametry
p_3=geom_point(data=NULL,aes(x=mu, y=0),alpha=0.9, colour="seagreen",size=3)

//znazorneni strednich hodnot pro data s variacnimi parametry
p_4=geom_point(data=NULL,aes(x=mu_N.[n_it,],alpha=0.1, y=0), colour="tomato",
  size=3)

//znazorneni hustoty pro data se skutecnymi parametry
p_5=geom_density(data=NULL,aes(x),alpha=0.2,color="seagreen",size=1)

//znazorneni hustoty pro data s variacnimi parametry
p_6=geom_density(data=NULL,aes(x_),alpha=0.2,color="tomato",size=1)

p_7=scale_x_continuous(breaks = scales::pretty_breaks(n = 15))
p_8=theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

p_1+p_2+p_4+p_3+p_5+p_6+p_7+p_8+labs(x = "data",y="hustota")

//vypsani variacnich parametru v posledni iteraci
print(mu_N.[n_it,])
print(k_N.[n_it,])
print(d_N.[n_it,])
print(a_N.[n_it])
print(b_N.[n_it])
print(n_it)

```

Výpis 2: Algoritmus počítající variační parametry pro n komponentní směs

```

//Algoritmus pocitajici variacni parametry tlakovych dat
library(ggplot2)
library(MCMCpack)
set.seed(5)

```

```

//pocet dat
N=nrow(h5)

//pocet komponent
J=2

//maximalni iterace
n_it=1000

//prah
eps=0.0001

//apriorni parametry
mu_=c(2.1,2.5)
k=c(1,1)
d=c(0.5,0.5)
a=1
b=1/400

//data - tlaky
x<-as.vector(unlist(log(h5[6])))

//parametry variacni aproximace
lambda<-matrix(nrow=n_it,ncol=2+3*J)
phi<-matrix(nrow=N,ncol=J)

theta<-matrix(nrow=N,ncol=J)

k_N.<-matrix(nrow=n_it,ncol=J)
mu_N.<-matrix(nrow=n_it,ncol=J)
d_N.<-matrix(nrow=n_it,ncol=J)
a_N.<-vector()
b_N.<-vector()

// parametry vychozi aproximace
k_N.[1,1:J]<-1
mu_N.[1,1:J]<-c(2,2.5)
d_N.[1,1:J]<-1

```

```

a_N.[1]<-1
b_N.[1]<-1/400

rozdil_mu_N.<-vector()
rozdil_d_N.<-vector()
rozdil_k_N.<-vector()

i<-2

podminka=0
while(podminka==0)
{
  for(j in 1:J)
  {
    for(n in 1:N)
    {
      //vypocet lokalnich variacnich parametru
      phi[n,j]=digamma(d_N.[i-1,j])-digamma(sum(d_N.[i-1,1:J]))
      -log(sqrt(2*pi))+(1/2)*(digamma(a_N.[i-1])-log(b_N.[i-1]))
      +((-a_N.[i-1]/(2*b_N.[i-1]))*x[n]^2)
      +(a_N.[i-1]/b_N.[i-1])*mu_N.[i-1,j]*x[n]
      -(1/2)*(1/k_N.[i-1,j]+(a_N.[i-1]/b_N.[i-1])*mu_N.[i-1,j]^2))

      theta[n,j]=exp(phi[n,j])
    }
  }

  //normalizace parametru theta
  theta=theta/rowSums(theta)

  //vypocet globalnich variacnich parametru
  for(j in 1:J)
  {
    lambda[i,2+j]=k[j]*mu_[j]+sum(x[1:N]*theta[1:N,j])
    lambda[i,2+j+J]=(-1/2)*(k[j]+sum(theta[1:N,j]))
    lambda[i,2+j+2*J]=d[j]+sum(theta[1:N,j])

    k_N.[i,j]=-2*lambda[i,2+j+J]
  }
}

```



```

mu_N.[i,j]=-(lambda[i,2+j])/(2*lambda[i,2+j+J])
d_N.[i,j]=lambda[i,2+j+2*J]

}

lambda[i,1]=a-1+J/2+N/2
lambda[i,2]=-b-1/2*(sum(x*x)+sum(mu_~2*k))

a_N.[i]=lambda[i,1]+1-J/2
b_N.[i]=(1/4)*sum(lambda[i,3:(2+J)]^2/lambda[i,(3+J):(2+2*J)])-lambda[i,2]

//ukoncuji kriterium - chyba normy rozdilu variacnich parametru v i a i-1
  iteraci
podminka=1
for(j in 1:J)
{
rozdil_mu_N.[j]<-sqrt(((sqrt((mu_N.[i,j])^2))-sqrt((mu_N.[i-1,j])^2)))^2)
rozdil_d_N.[j]<-sqrt(((sqrt((d_N.[i,j])^2))-sqrt((d_N.[i-1,j])^2)))^2)
rozdil_k_N.[j]<-sqrt(((sqrt((k_N.[i,j])^2))-sqrt((k_N.[i-1,j])^2)))^2)
rozdil_b_N.<-sqrt(((sqrt((k_N.[i])^2))-sqrt((k_N.[i-1])^2)))^2)

if((rozdil_mu_N.[j])>eps||(rozdil_d_N.[j])>eps||(rozdil_k_N.[j])>eps||
(rozdil_b_N.)>eps)
{
podminka=0
break
}
}
i<-i+1
}

//aktualni maximalni iterace
n_it<- i-1

//znazorneni variacnich parametru v dane iteraci
for (j_it in 1:J)
{
p1=ggplot(data=NULL,aes(x =m<-1:(n_it) , y = mu_N.[(1:n_it),j_it]))
+geom_point(color = 'green')+geom_line(color="grey")

```

```

+labs(x="pocet iteraci")+labs(y=paste("mu_N.",j_it))
+ scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

p2=ggplot(data=NULL,aes(x = m<-1:(n_it), y = d_N.[(1:n_it),j_it]))
+geom_point(color = 'green') + geom_line(color="grey")
+labs(x="pocet iteraci")+ labs(y=paste("d_N.",j_it))
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

p3=ggplot(data=NULL,aes(x = m<-1:(n_it), y = k_N.[(1:n_it),j_it]))
+geom_point(color = 'green') + geom_line(color="grey")
+labs(x="pocet iteraci")+labs(y=paste("k_N.",j_it))
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

show(p1)
show(p2)
show(p3)
}

ggplot(data=NULL,aes(x = m<-1:(n_it), y = a_N.))
+geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

ggplot(data=NULL,aes(x = m<-1:(n_it), y = b_N.))
+geom_point(color = 'green')
+geom_line(color="grey")+labs(x="pocet iteraci")
+scale_x_continuous(breaks = scales::pretty_breaks(n = 10))
+scale_y_continuous(breaks = scales::pretty_breaks(n = 15))
+theme(axis.text=element_text(size=12),axis.title=element_text(size=14))

//znazorneni histogramu realnych dat

```

```

ggplot(data=NULL,aes(x,y=..density..))
+geom_histogram(color="seagreen", fill="seagreen3",bins = 25)
+scale_x_continuous(breaks = scales::pretty_breaks(n = 15))

//generovani dat s variacnimi parametry
x_<-vector()
sd_ = vector()
for(s in 1:1000)
{
tau_=rgamma(1, shape=a_N.[n_it],rate=b_N.[n_it])
for(j in 1:J)
{
sd_[j] <- 1/sqrt(k_N.[n_it,j]*tau_)
}
alpha_=rdirichlet(1, d_N.[n_it,])
muP=rnorm(J, mu_N.[n_it,], sd_)

components_<-sample(1:J,prob=alpha_,size=N,replace=TRUE)
x_<-c(x_, rnorm(n=N,mean=muP[components_],tau_*(-0.5)))

}

p_1=ggplot()

//znazorneni histogramu pro realna data
p_2=geom_histogram(data=NULL,aes(x,y=..density..),alpha=0.3,color="seagreen4",
fill="seagreen",bins = 25)

//znazorneni strednich hodnot pro data s variacnimi parametry
p_3=geom_point(data=NULL,aes(x=mu_N.[n_it,], y=0), colour="tomato",size=3)

//znazorneni hustoty pro realna data
p_4=geom_density(data=NULL,aes(x),color="seagreen",size=1)

//znazorneni hustoty pro data s variacnimi parametry
p_5=geom_density(data=NULL,aes(x_),color="tomato",size=1)

p_6=scale_x_continuous(breaks = scales::pretty_breaks(n = 15))

```

```

p_7=theme(axis.text=element_text(size=11),axis.title=element_text(size=14))

p_1+p_2+p_3+p_4+p_5+p_6+p_7+labs(x = "data",y="hustota")

//vypsani variacnich parametru v posledni iteraci
print(mu_N.[n_it,])
print(k_N.[n_it,])
print(d_N.[n_it,])
print(a_N.[n_it])
print(b_N.[n_it])
print(n_it)

```

Výpis 3: Algoritmus počítající variační parametry tlakových dat